The Hahn–Banach theorem: the life and times

Lawrence Narici \textsuperscript{a,\ast}, Edward Beckenstein \textsuperscript{b,1}

\textsuperscript{a} Department of Mathematics, St. John's University, Jamaica, NY 11439, USA
\textsuperscript{b} Department of Mathematics, St. John's University, Staten Island, NY 10301, USA

Received 7 May 1995; revised 5 January 1996

Abstract

Without the Hahn–Banach theorem, functional analysis would be very different from the structure we know today. Among other things, it has proved to be a very appropriate form of the Axiom of Choice for the analyst. (It is not equivalent to the Axiom of Choice, incidentally; it follows from the ultrafilter theorem which is strictly weaker.) Riesz and Helly obtained forerunners of the theorem in the turbulent mathematical world of the early 1900s. Hahn and Banach independently proved the theorem for the real case in the 1920s. Then there was Murray's extension to the complex case—easy, once you realize that \( f(x) = \text{Re} f(x) - i\text{Re} f(ix) \). Can continuous linear maps be extended as easily as linear functionals? Banach and Mazur had already proved that they could not in 1933 but it was not until Nachbin's 1950 result that a definitive answer was achieved to this more general question. In this article, we discuss the mathematical world into which the theorem entered, examine its connection to the axiom of choice, look at some ancestors, mention some of its consequences and consider some of its principal variations.© 1997 Elsevier Science B.V.

Keywords: Hahn–Banach theorem

1. What is it?

In its elegance and power, the Hahn–Banach theorem is a favorite of almost every analyst. Some of its sobriquets include The Analyst’s Form of the Axiom of Choice and The Crown Jewel of Functional Analysis. Its principal formulations are as a dominated extension theorem and as a separation theorem. As the paterfamilias of the extension version, let us take the following:

Let \( M \) be a subspace of a linear space \( X \) over \( \mathbb{R} \), let \( p \) be a sublinear (i.e., subadditive and positive homogeneous) functional defined on \( X \) and let \( f \) be a linear form defined

\{
* Corresponding author. E-mail: narici1@svmu.sjohns.edu.
1 E-mail: beckenst@svmu.sjohns.edu.
0166-8641/97/$17.00 © 1997 Elsevier Science B.V. All rights reserved.
PII S0166-8641(96)00142-3
The theorem asserts the existence of a linear extension $F$ of $f$ to all of $X$ such that $F$ is dominated by $p$ everywhere.

$$F: X \quad \quad F \leq p$$

$$f: M \rightarrow \mathbb{R} \quad f \leq p$$

2. Why is it important?

The Hahn--Banach theorem is a powerful existence theorem whose form is particularly appropriate to applications in linear problems. Some of the ways in which it resonates throughout functional analysis include:

- duality theory;
- Cauchy integral theorem for vector-valued analytic functions $x: D \rightarrow X$, $X$ a Banach space, $D$ a domain of the complex plane $\mathbb{C}$ (Narici and Beckenstein [58, p. 162]);
- Helly's criterion for solving systems of linear equations in reflexive normed spaces (see Section 5 as well as Narici and Beckenstein [58]).

Its reach extends beyond functional analysis to:

- proof of the existence of Green's functions (Garabedian and Schiffman [23]);
- Banach's solution of the 'easy' problem of measure (Bachman and Narici [1, p. 188f]);
- applications to control theory (Leigh [48], Rolewicz [75]);
- applications to convex programming (Balakrishnan [2]);
- applications to game theory (König [46]);
- a formulation of thermodynamics (Feinberg and Lavine [21]).

3. A short history of analysis

In the nineteenth century, 'vector' meant 'n-tuple'. Toward the end of the century, its scope was extended to include 'sequence'—for some, anyway. There were only fleeting contacts between geometric ideas and analysis for the most part and notions of proof were quite relaxed, to say the least. The geometric theorem-proof style, common today in most areas of mathematics, had to wait for the insights of Peano and Hilbert & Co. To 'prove' something, you merely stated your case and argued its plausibility. It was unfortunately similar to the rash manner in which the social 'sciences' provide 'proofs' in the modern era. We briefly illustrate how cavalier even such greats as Fourier and Euler were in this regard in Section 3.3.

In the period 1890–1915 notions of structure were emerging in analysis and geometric perspectives were being adopted. Standards of rigor were greatly improved and new integrals made it possible to unify several different things.
3.1. Structure

Mathematics had matured to the point where the similarities between manipulating different concrete objects were becoming apparent. A way was needed to be able to express this indifference to actual identity. The ultimate framework was to let the objects be points of an arbitrary set whose interactions were governed by a set of rules. It happened first in algebra. There, Peano [62] in 1888 defined vector space and linear map axiomatically. No more were vectors n-tuples or sequences; now you could not know exactly what the 'vectors' were. Significantly, this opened the way to vector spaces of arbitrary dimension, in particular to function spaces. But even though Pincherle wrote a book [65] about linear spaces in 1901, Peano's idea was mostly ignored. Still, the idea of defining a space abstractly as 'objects' that obeyed certain rules was one whose time had come. Groups (a term coined by Galois) were defined on an arbitrary set for the first time by Weber [86] in 1895; field in 1903.

In analysis it took a little longer than it did in algebra for the idea of structure to take hold. The concrete objects here were functions but confusion persisted about exactly what a function was. Dirichlet (1837) defined a numerical-valued function of a real variable to be a table, or correspondence or correlation between two sets of numbers. Riemann (1854) saw problems with the intuitive notion of function. To make the point that our understanding was too primitive, he invented a function—defined by a trigonometric series—which is continuous for irrational values of the independent variable, discontinuous for rational values. Weierstrass's (1874) classic example of a nowhere differentiable, continuous function made the point even more dramatically. As a result of these discoveries, Dedekind, Weierstrass, Méré and Cantor, by different routes, made the \( \varepsilon-\delta \) technique part of the standard répertoire of analysis.

Pincherle insisted on distinguishing between the function and the values it assumed. He said that mathematicians should use \( f \) rather than \( f(x) \), to think of the function itself as an entity, divorced from its values. He and others decried the confusion between a linear map and the matrix which represented it in a particular coordinate system, a problem that is unfortunately still with us. Concomitant with the point of view that functions were entities in themselves, Volterra [85] in 1888 suggested that we should be thinking of functions defined on new domains such as on all continuous curves in a square, and doing analysis on them—no easy trick without general topology at one's disposal. He called these new kinds of functions fonctions de ligne, the ligne being the continuous curve within the square.

But what is a curve? protested Peano. The term meant something like a continuous image of \([0, 1]\) in the unit square. Peano's space-filling curve eloquently demonstrated the diverse possibilities that such a definition permitted. Hadamard was intrigued by Volterra's suggestion, however, and persisted. In 1903 he called the new functions of functions functionals, analysis on them functional analysis. Part of this was not new. In the early 1800s there was also consideration of functions whose domains were functions—derivatives, Laplace transforms, shift operators—but the radical thing at that time was applying algebraic rules to them, a notion heretofore thought only to apply to numbers. The time had now come to consider the analytic properties of such operators.
Fréchet [17] in 1904 propounded ideas of limit and continuity in sets which did not consist of numbers. In his 1906 thesis [19] he defined the present notion of metric (He did not coin the term metric space, incidentally. Hausdorff [26] introduced the more geometric-sounding nomenclature in 1913.) and investigated concrete metric spaces in which the 'points' were functions. He saw and stressed the importance of completeness, compactness and separability.

### 3.2. Point of view—geometric perspective

Geometry had been ‘algebraized’ in the early seventeenth century by Descartes and Fermat. It was time for geometry’s revenge in the late nineteenth and early twentieth, time for it to ‘geometrize’ analysis. Schmidt [77] and Fréchet [20] in 1908 introduced the language of geometry into the Hilbert space $\ell_2$, first spoke of the norm (in its present notation $\|x\|$) and of the triangle inequality for the norm. In 1913 Riesz [73] described the solution of systems of homogeneous equations

$$f_i(x) = a_{i1}x_1 + \cdots + a_{in}x_n = 0, \quad 1 \leq i \leq n,$$

as an attempt to find $x = (x_1, \ldots, x_n)$ orthogonal to the linear span $[f_1, \ldots, f_n]$ where $f_i = (a_{i1}, \ldots, a_{in})$, i.e., he viewed solving the equations as an attempt to identify the orthogonal complement of the linear span $[f_1, \ldots, f_n]$ of the $f_1, \ldots, f_n$. Significantly, the 'equations', the $f_i$, achieved vector status and stood on equal footing with the 'variables'. Hilbert and his school also spoke of orthogonal expansions. Helly and others, relying on earlier work [53] of Minkowski (1896) introduced ideas about convexity into the bloodstream of analysis. The legacy of those ideas is still very much with us.

### 3.3. Precision

Two principal defects of analysis in the seventeenth century were its capricious intuitiveness and its purely formal manipulation of symbols. As an example of this intuitiveness, consider Johann Bernoulli’s (1693) mystic dogma that ‘a quantity which is increased or decreased by an infinitely small quantity is neither increased nor decreased’. As Bishop Berkeley furiously pointed cut in The Analyst in 1734, this gave analysts the best of both worlds: they could treat this schizophrenic ‘ghost of a departed quantity’ as something until the last step of an argument and then jettison it as nothing. Nowadays, some applied mathematicians retain ‘the little zero’ $dx$ but discard ‘higher order’ terms $dx^2$, $dx^3$, etc., at moments apparently determined more by convenience than rigor.

For pure manipulation of symbols in series and products without regard to convergence, the master was Euler. Consider his ‘proof’ that

$$e^x = \sum_{n \geq 0} \frac{x^n}{n!}$$

by means of taking the ‘limit’ as $n \to \infty$ in the binomial expansion

$$\left(1 + \frac{x}{n}\right)^n = 1 + x + \frac{n(n-1)}{2!} \frac{x^2}{n^2} + \frac{n(n-1)(n-2)}{3!} \frac{x^3}{n^3} + \cdots.$$
This apparently did not perturb his mathematical conscience. Despite Lagrange’s protests, Fourier was equally uninhibited in his 1822 classic on heat, La Théorie Analytique de la Chaleur. Having developed an expansion of a certain function in a series of sines and cosines, he says ‘We can extend the same results to any functions, even to those which are discontinuous and entirely arbitrary’. He formally manipulates symbols, leaving convergence to take care of itself, and obtains an expansion of an ‘arbitrary’ odd function in a sine series.

Though the influence of the work of Cauchy, Riemann and Weierstrass had already raised standards. The work of Hilbert and his school on the foundations of geometry elevated the standards of rigor so much that most earlier mathematical work looks shabby by comparison.

3.4. New tools: the new integrals

Considerable effort was expended in the nineteenth century on the problem of solution of systems of infinitely many equations in infinitely many unknowns. (Try and get mathematicians not to try to solve equations!) In the linear case the simultaneous linear equation problem could be stated: given linear functionals $f_i$ and scalars $c_i$, find $x$ such that $f_i(x) = c_i$. However many $f$'s (and $c$'s) there were, that was the number of coordinates $x$ was supposed to have. When there are infinitely many $f$'s and $c$'s, $x$ must have infinitely many components or coordinates—must be a sequence, that is, rather than a tuple. Considerable progress in solving infinite systems of linear equations was achieved by cleverly generalizing determinants. The basic idea was to truncate the infinite system of linear equations and then take a limit. A serious weakness of the approach was its dependence on infinite products which converge only under highly restrictive circumstances. Lebesgue and Stieltjes’ new theories of the integral made it possible to unify the problems, of which the following are two special cases.

(1) Fourier series. Given a sequence $(g_n)$ of cosines, say, and $(a_n)$ of numbers, perforce from $l_2$, find a function $x$ for which these were the Fourier coefficients, i.e., such that $\int x(t)g_n(t)\,dt = a_n$ for every $n \in \mathbb{N}$. Is $x$ unique?

(2) Moment problems. Given a sequence $(a_n)$ of numbers, find a function $x$ such that $\int t^n x(t)\,dt = a_n$ for every $n \in \mathbb{N}$.

4. What Riesz did

Borrowing some things already done in Hilbert space, Riesz [71,72] (1910–1911) set out to solve the following problem: for $p > 1$ (so he could use the Hölder and Minkowski inequalities which he had just generalized):

(P) Given infinitely many $y_\alpha$ in $L_q[a,b]$ and scalars $c_\alpha$, find $x$ in $L_p[a,b]$ such that

$$\int_a^b x(t)y_\alpha(t)\,dt = c_\alpha.$$
His solution and method of attack bore no resemblance to what had come before. For there to be such an $x$, he showed that the following necessary and sufficient connection between the $y$'s and the $c$'s had to prevail:

\begin{equation}
(*) \text{ for any finite set of indices } s \text{ and any scalars } a_s \text{ there should exist } K > 0 \text{ such that }
\left| \sum a_sc_s \right| \leq K \left( \int_a^b \left| \sum a_sy_s \right|^q \right)^{1/q}.
\end{equation}

Note that $(*)$ implies that if $\sum a_sy_s = 0$ then $\sum a_sc_s = 0$ as well. So, if we define a linear functional $f$ on the linear span $M$ of the $y$'s in $L_q[a,b]$ by taking $f(y_s) = c_s$, the $f$ so obtained is well-defined. Not only that, for any $y$ in $M$, $|f(y)| \leq K\|y\|_q$ so in today’s language, we would say that $f$ is bounded or continuous on $M$. If there is an $x$ in $L_p$ which solves (P), then he showed that $f$ has a continuous extension to the whole space. The ability to solve linear equations, in other words, implies being able to continuously extend a bounded linear functional to the whole space. Thus, Riesz’s solution to (P) constitutes a special case of the Hahn–Banach theorem.

Riesz then changed spaces and turned to the following variant of the problem:

(Q) Given $y_s \in C[a,b]$, and scalars $c_s$, find $x \in BV[a,b]$ (bounded variation) such that
\[
\int_a^b y_s(t) dx(t) = c_s.
\]

Adapting his earlier methods, he solved it with a condition that looked very much like the boundedness condition $(*)$. He realized the importance of the condition and proved that any ‘continuous additive’ map satisfied such a condition and conversely, where by ‘continuous’ he meant sequentially continuous with respect to the sup norm. In each case he proved a special case of the Hahn–Banach theorem and identified the continuous dual of a normed space.

5. Enter Helly

Riesz did not view things in terms of defining and extending continuous linear forms. Banach in 1923 [4], however, solved the problem of measure by using transfinite induction to extend nonnegative linear functionals. Indeed (Saccoman [76]), Banach’s argument implies the following special case of the Krein–Rutman [3,7] extension theorem.

**Theorem.** Let $M$ be a linear subspace of an ordered vector space $X$ with order unit $e \in M$ and $f$ a nonnegative linear functional defined on $M$. Then there exists a nonnegative linear functional $F$ defined on $X$ such that $F(x) = f(x)$ on $M$. 

Helly in 1912 [31] viewed things in terms of extending continuous linear forms and gave the precursor to the argument that Hahn in 1927 [28] and Banach in 1929 [5] each used later to prove the Hahn–Banach theorem—namely, by reducing the problem to showing that a continuous linear form defined on a subspace $M$ of a normed space can be extended to an enlargement by one vector to $[M \cup \{x\}]$ without increasing its norm. He revisited (Q) and gave a different proof nine years later (1921)—he had been a prisoner of war in Russia as a soldier in the Austrian army in the meantime. Instead of particular spaces $\ell_p$, $L_p[a,b]$, and $C[0,1]$, he deals with a general norm (though he did not call it that, nor use the notation $\|x\|$) on a general sequence space—specifically, any vector subspace of $\mathbb{C}^\mathbb{N}$. This, of course, covered the $\ell_p$ spaces and many others such as $L_2$ which could be identified with $\ell_2$. Helly linked his general norm with some of Minkowski’s earlier ideas concerning convexity. Minkowski had already observed the correspondence between ‘norms’ on a subspace of $\mathbb{R}^n$ and ‘symmetric convex bodies’ (closed, symmetric, bounded, convex sets which have 0 as an interior point), an idea which reemerged decades later when locally convex spaces were developed.

Given a normed subspace $X$ of $\mathbb{C}^\mathbb{N}$, Helly considered the subspace

$$X' = \left\{ (u_n) \in \mathbb{C}^\mathbb{N} : \sum_{n \in \mathbb{N}} x_n u_n < \infty \right\}$$

for all $(x_n) \in X$, i.e., $(u_n)$ such that $(u_n x_n)$ is summable for all $(x_n) \in X$. For example, if $X = c$ or $c_0$, then $X' = \ell_1$; if $X = \ell_1$, then $X' = \ell_\infty$; of course, you do not always get the continuous dual of $X$ this way—you do not if you take $X = \ell_\infty$, for example. Anyway, for $x = (x_n) \in X$ and $u = (u_n) \in X'$, Helly defined a linear form on $X$ (bilinear form on $X \times X'$, making $(X, X')$ a dual pair) by taking

$$\langle x, u \rangle = \sum_{n \in \mathbb{N}} x_n u_n.$$

Using an idea of Minkowski’s, he normed $X'$ by taking

$$\|u\| = \sup \left\{ \frac{|\langle x, u \rangle|}{\|x\|} : x \neq 0 \right\}.$$

The dual norm on $X$ obtained by this technique yields the original norm on $X$. Nowadays such pairs with absolute convergence of $\sum x_n u_n$ are called Köthe sequence spaces and Köthe duals. By the Cauchy-Schwarz inequality, $|\langle x, u \rangle| \leq \|x\| \|u\|$, so the linear functionals obtained in this manner are continuous or bounded (beschränkt) as Riesz called them.

Helly then set out to solve

(R) Given $u_i \in X'$, $(c_i) \in \mathbb{C}^\mathbb{N}$, find $x \in X$ such that

$$\langle x, u_i \rangle = c_i \quad \text{for each } i \in \mathbb{N}.$$

He split the problem into two parts:
(A) find a linear map \( f : X' \to \mathbb{C} \) such that \(|f(u)| \leq k\|u\|\) for some \( k > 0 \) and all \( u \) in \( X' \) with \( f(u_i) = e_i \); and

(B) once \( f \) has been found (if it can be found), find \( x \in X \) such that \( \langle x, u \rangle = f(u) \) for all \( u \in X' \).

Helly solved (A) by induction and a result of his on convex sets; he discovered that the \( x \) of (B) could not always be found. He (and Riesz) thus became the first to exhibit nonreflexive Banach spaces.

In summary, Helly's principal contributions were the following:

- defined and worked with a general sequence space endowed with a general norm;
- utilized various notions about convexity;
- introduced the rudiments of duality theory;
- realized the generality of Riesz's continuity condition (*) and defined the infimum of the \( K \) that satisfy (*) as the Maximalzahl, i.e., what we now call the norm of the linear functional.

6. Hahn and Banach

Hahn [28] and Banach [5] took an even more general approach. Even though both used the same technique that Helly used—reducing the problem to the case of enlarging the domain of the functional by just one vector—neither credited Helly with the central idea for the proof of the Hahn–Banach theorem. Banach, however, referred to Helly's 1912 paper in deriving as his first application of the theorem the result of Riesz that Helly had proved. Aside from that, Hahn and Banach went a long way to shaping functional analysis as we know it today.

- They defined the general normed space. Hahn [27] in 1922 and Banach [3] in 1923 did it independently. Each of them required completeness. Banach [6] later removed it in his book, distinguishing between normed and Banach spaces. (The general notion of norm was 'in the air' at this time. Wiener, too, in 1922 [87] defined it contemporaneously.)
- They abandoned systems of linear equations and considered the general problem of extending a continuous linear form defined on a general normed space, not a sequence space as Helly had done. Thus, they formulated the theorem as we know it today.
- They defined the dual space of a general complete normed space and proved that it too is a complete normed space with respect to the standard norm.
- They defined reflexivity and realized that a normed space \( X \) is generally embedded in its second dual \( X'' \).
- They used transfinite induction (Helly had used ordinary induction). The way it was used here became an essential tool of the analyst from that time forward.

In 1927 Hahn [28] returned to Helly's 1921 results [32] in the context of general real Banach spaces. His proof of Helly's results by transfinite induction instead of ordinary induction simplified and generalized them. Although transfinite induction had been
used by analysts before, with the exception of Banach's treatment [4] of the problem of measure, it had not been employed like this. Hahn, of course, did not use the Zorn's lemma formulation of transfinite induction, for that did not exist until 1935, but rather used ordinals. Aside from treating the earlier problem strictly as one of extending linear functionals, Hahn also formally introduced the notion of dual space (polare Raum) for the first time, noted that $X$ is embedded in its second dual $X''$ and defined reflexivity (regularität). Duality theory had reached adolescence.

Unaware of Hahn's work, Banach [5] also used well-ordering and transfinite induction to prove the Hahn–Banach theorem in 1929. He acknowledged Hahn's priority in his book and generalized the result slightly; instead of considering the linear form $f$ to be dominated by a multiple of the norm, he considered an $f$ dominated by a sublinear functional; he made no other use of the greater generality, however. Nobody did until locally convex spaces had been introduced. Then Banach's more general result was quite useful.

Their work has the following immediate consequences:

- **Norm-preserving extensions.** Given a continuous linear functional $f$ defined on a subspace of a normed space, there exists a continuous linear extension $F$ defined on the whole space such that $\|f\| = \|F\|$.

- **Nontrivial continuous linear forms.** A linear form $f$ on a locally convex space $X$ is continuous if and only if there is a continuous seminorm $p$ on $X$ such that $|f| < p$. Moreover, if $X$ is Hausdorff, and $x \neq 0$, there must be a continuous seminorm $p$ on $X$ such that $p(x) \neq 0$. The Hahn–Banach theorem implies that, for any nonzero vector $x$, there is a continuous linear functional $f$ on $X$ such that $f(x) = p(x) \neq 0$. Consequently, if every continuous linear functional vanishes on a vector $x$, then $x = 0$.

7. Uniqueness

In the standard proof (i.e., Banach's) of the lemma to the Hahn–Banach theorem, the one in which it is shown that a dominated extension of the same norm exists on the linear subspace $[M \cup \{x\}]$ for $x \notin M$, a number $c$ is chosen arbitrarily between two others. Herein lies the nonuniqueness of the extension. Taylor [84] and Foguel [16] characterized the normed spaces $X$ for which each continuous linear functional on any subspace of $X$ has a unique linear extension of the same norm: they are those $X$ with a strictly convex dual. If we focus on just one subspace $M$ of $X$ then continuous linear forms on $M$ have unique extensions of the same norm if and only if the annihilator $M^\perp$ of $M$ has unique best approximations in $X'$, i.e.,

Phelps [64]: If $M$ is a linear subspace of the normed space $X$ then $f \in M'$ (the continuous dual of $M$) has a unique extension of the same norm in $X'$ if and only if for each $g \in X'$ there exists a unique $h \in M^\perp = \{u \in X': u|_M = 0\}$ such that

$$\|g - h\| = \inf \{\|g - u\|: u \in M^\perp\}.$$  

This result is generalized by Park [61].
8. The Axiom of Choice

A few words on the Axiom of Choice (AC) are in order as most proofs of the Hahn-Banach theorem use its Zorn’s lemma variant. There are some notable exceptions, however. In their 1968 textbook, Garnir, de Wilde and Schmets [24] use only the Axiom of Dependent Choices (ADC). Given a nonempty set $X$ and $R \subseteq X \times Y$ such that, for every $x \in X$, \( \{ y \in X : (x, y) \in R \} \neq \emptyset \), then for every $w \in X$ there exists a sequence $(x_n)$ from $X$ such that $x_1 = w$ and $(x_n, x_{n+1}) \in R$ for every $n \in \mathbb{N}$.

to prove a Hahn–Banach theorem for separable spaces. (Garnir et al. [24] claim that they only use the Countable Axiom of Choice but Blair [8] shows that they really need ADC.) ADC is weaker than AC but implies the Countable Axiom of Choice. ADC, incidentally, is strong enough to prove Urysohn’s lemma as well as the Baire Category Theorem [8]. Another ‘constructive’ version is that of Ishihara [38].

A nonphilosophical objection to using AC to prove the Hahn–Banach theorem is that the arbitrariness of the functional so obtained limits the information which may be gleaned from it. Mulvey and Pelletier [54] consider a context in which dependence on AC may be circumvented. Locales generalize the lattice of open sets of a space without reference to the points of the space. Mulvey and Pelletier [54] systematically use locales to prove a version of the Hahn–Banach theorem in any Grothendieck topos.

Does the Hahn–Banach theorem (HB) imply the Axiom of Choice, as Tihonov’s theorem does? As is well known, the Axiom of Choice implies the Ultrafilter Theorem (UT), that every filter is contained in an ultrafilter. (UT, incidentally, is equivalent to the existence of the Stone–Čech compactification of any Tihonov space.) Halpern [29] proved that the Ultrafilter Theorem does not imply the Axiom of Choice. Łoś and Ryll-Nardzewski [49] and Luxemburg [50–52] proved that the Ultrafilter Theorem sufficed to prove the Hahn–Banach theorem. Pincus [66,67] proved that the Hahn–Banach theorem does not imply the Ultrafilter Theorem. We therefore have the following irreversible hierarchy: $\text{AC} \Rightarrow \text{UT} \Rightarrow \text{HB}$.

9. The complex case

The complex version of the theorem hinged on the discovery of the intimate relationship between the real and complex parts of a complex linear functional $f$, namely that

$$\text{Re } f(x) = -i \text{Im } f(ix).$$

By thus reducing the complex case to the real case, the complex version was first proved by F. Murray [55] in 1936 for $L_p[a, b]$, $p > 1$. His method, however, was perfectly general and was used (and acknowledged) by Bohnenblust and Sobczyk [10] in 1938 who proved it for arbitrary complex normed spaces. They were the first to refer to the theorem as the Hahn–Banach theorem, incidentally. Also, by reducing things to the real case, Soukhomlinov [82] in 1938 and Ono in 1953 [59] obtained the theorem for vector
spaces over the complex numbers and the quaternions. In contrast with the methods of reduction to the real case, Holbrook in 1975 [33] proved it in a way that does not depend on the choice of the Archimedean-valued scalar field, be it \( \mathbb{R} \), \( \mathbb{C} \), or the quaternions. He used Nachbin’s approach, discussed in Section 10.1, together with an “intersection property” (Holbrook’s Lemma 1) shared by all three of the fields.

10. Related questions

10.1. The range side

The success of the Hahn–Banach theorem suggested the investigation of questions of continuous extendibility of more general continuous linear maps. One variation is to replace the field \( \mathbb{R} \) or \( \mathbb{C} \) by a normed space \( Y \). For real normed spaces \( X \) and \( Y \), let \( A \) be a continuous linear map of a subspace \( M \) of \( X \) into \( Y \). Find a continuous linear extension \( \mathcal{A} \) of \( A \) to \( \mathcal{A}/M \) with the same norm. Say that \( Y \) is extendible if for any subspace \( M \) of any real normed space \( X \), such an \( \mathcal{A} \) exists.

\[
\mathcal{A}: \quad X \quad \quad \quad \| \mathcal{A}x \| \leq k \| x \|
\]

\[
\| \mathcal{A} \| = \| A \|
\]

\[
A: \quad M \rightarrow Y \quad \| Ax \| \leq k \| x \|
\]

Banach and Mazur [7] quickly demonstrated that there are instances where there is no such \( \mathcal{A} \). But consider the special case \( Y = \mathbb{R}^n, n > 1 \), with any of the norms \( \| \cdot \|_p, 1 \leq p \leq \infty \). Even though the topology is the same in every case, \((\mathbb{R}^n, \| \cdot \|_\infty)\) is extendible while none of the others \((\mathbb{R}^n, \| \cdot \|_p), 1 \leq p < \infty\), is. As Nachbin [56] and Goodner [25] discovered, a real normed space \( Y \) is extendible iff it has

The binary intersection property. Every collection of mutually intersecting closed balls has nonempty intersection (Nachbin [56], Goodner [25], Kelley [40]; cf. Narici and Beckenstein [58]).

Examples on extendible spaces

(a) The Euclidean normed space \( \mathbb{R}^2 \) does not have the binary intersection property for one can draw three mutually intersecting circles whose intersection is empty; for essentially the same reason, neither does \( \mathbb{R}^n \) with any of the norms \( \| \cdot \|_p, 1 \leq p < \infty \).

(b) \( B(T, \mathbb{R}) \), the space of bounded functions on any set \( T \) with sup norm has the binary intersection property. One may take \( T = \mathbb{N} \) to get \( \ell_\infty \), or \( T = \{1, 2, \ldots, n\} \) to get \((\mathbb{R}^n, \| \cdot \|_\infty)\). Even though \( \ell_\infty \) has the binary intersection property, the closed subspace \( c_0 \) of \( \ell_\infty \) of null sequences does not, so the binary intersection property is not a hereditary property. This example of \( c_0 \) also shows that sup norms, despite the ‘cubic’ nature of the balls they produce, do not always produce extendible spaces.

(c) Consider the linear space \( C(T, \mathbb{R}) \) of continuous real-valued maps on the compact Hausdorff space \( T \) with sup norm. If \( T \) is extremally disconnected (disjoint open sets have disjoint closures or, equivalently, open sets have open closures) then \( C(T, \mathbb{R}) \) has
the binary intersection property. Extremal disconnectedness, incidentally, is the notion
Stone introduced in proving that every complete Boolean algebra is Boolean algebra iso-
morphic to the Boolean algebra of clopen subsets of an extremally disconnected compact
Hausdorff space. As the power set of any set is a complete boolean algebra, there are
clearly plenty of extremally disconnected spaces.

(d) Let us leave normed spaces for a moment. Let $X$ be a locally convex space over
$K = \mathbb{R}$ or $\mathbb{C}$, let $S$ be any set and let $\mathbb{K}^S$ carry the Tihonov topology. Any continuous
linear map $A$ defined on any subspace $M$ of $X$ into the product may be continuously
extended as a linear map to all of $X$. As infinite products of normed spaces (Narici and
Beckenstein [58, 7.4.5, p. 137] are never normable, this is a different sort of result. For
finite $S$, note that the Tihonov topology is the sup norm topology.

An extendible space must be a Banach space, because it must be possible to extend
the identity map $1 : Y \to Y$ to $\bar{1}$ on the norm-completion $\bar{Y}$ of $Y$. If $(y_n)$ is Cauchy
in $Y$, it is convergent, to $y \in \bar{Y}$. As $\bar{1}$ is continuous, $\bar{1}y_n = y_n \to \bar{1}y \in Y$ (the range of
$\bar{1}$ is the same as that of $1$, $Y$, by the definition of extendibility).

Another quality that extendible spaces $Y$ must have is that of projectibility: if $X$ is
any real normed space that contains $Y$ then there must be a continuous projection $E$
of $X$ onto $Y$ of norm 1. Equivalently, $Y$ is topologically complemented in each space
in which it is norm-embedded. As there is no continuous projection from $\ell_\infty$ onto $c_0$
(Narici and Beckenstein [58, Example 5.8-1]), $c_0$ is not extendible.

The principal words on extendibility of real Banach spaces are contained in the fol-
lowing result:

Nachbin [56], Goodner [25] and Kelley [40]: For a real normed space $Y$, the following
are equivalent:

(a) $Y$ is extendible.
(b) $Y$ is projectible.
(c) $Y$ has the binary intersection property.
(d) $Y = C(T, \mathbb{R})$ with sup norm, where $T$ is a compact extremally disconnected
Hausdorff space.
(e) $Y$ is a complete Archimedean ordered vector lattice with order unit.

The complex case

The binary intersection property fails to characterize extendibility for complex spaces.
For example, $\mathbb{C}$ is extendible but does not have the binary intersection property. Har-
sumi [30] showed the equivalence of (a) and (d) for complex spaces. He showed that a
complex normed space $Y$ is extendible iff $Y$ is norm-isomorphic to $C(T, \mathbb{C})$ where $T$ is
a compact Hausdorff extremally disconnected space.

10.2. The domain side

Consider the problem of identifying those normed spaces $X$ that have the property
that any continuous linear map $A$ of any subspace $M$ into any normed space $Y$ has a
linear extension with same norm.
The problem was solved by Kakutani [39] in 1941 in the real case and Bohnenblust [9] in the complex case. The $X$ for which this is true are those $X$ for which $\dim X \leq 2$ or $X$ a Hilbert space.

10.3. Superspaces

Single out a Banach space $M$. For what $M$ does every superspace $X$ of $M$ and every continuous linear map $A$ of $M$ into any normed space $Y$ have a continuous extension to $X$?

$$\begin{align*}
\overline{A}: & \quad X \\
\mid & \\
A: & \quad M \to Y \\
\end{align*}$$

$\|A\| = \|A\|$, $X$ fixed

$\|Ax\| \leq k\|x\|$

It turns out that the class of such $M$ is the class of extendible spaces.

11. Minimal sublinear functionals

An interesting alternate approach to the Hahn-Banach theorem was developed by König [41–44,46], Fuchsteiner and König [22], and Simons [78–80]. Not only does it provide a different proof of the Hahn-Banach theorem, but it also points the way to more general theorems of Hahn–Banach type. An outline of the method follows.

Sublinear Functionals. For any vector space $X$, a subadditive, positive-homogeneous map $p: X \to \mathbb{R}$ is a sublinear functional. $X^\ast$ denotes the class of all sublinear functionals on $X$.

We order $X^\ast$ pointwise: $p \leq q$ if and only if $p(x) \leq q(x)$ for all $x \in X$. A minimal sublinear functional is a minimal element of $(X^\ast, \leq)$.

• For a real vector space $X$, a sublinear functional $p$ on $X$ is linear if and only if it is minimal.

König et al. reverse the usual Hahn–Banach theorem proofs in that the search is not for a maximal extension but for a minimal sublinear functional. The arguments follow the following sequence for a real vector space $X$.

1. For any $p \in X^\ast$ there exists a linear functional $h$ on $X$ such that $h \leq p$.
2. Given any linear form $f$ defined on a subspace $M$ of $X$ with $f \leq p$ on $M$ there exists $q \in X^\ast$ such that $q \leq f$ on $M$ and $q \leq p$ on $X$.
3. By (1), there exists a linear form $F$ defined on $X$ such that $F \leq q$, where $q$ is as in (2).
(4) Since $F \leq q \leq f$ on $M$ it follows (from the minimality of $f$) that $F = f$ on $M$. $F$ is therefore the desired dominated extension (dominated by $p$) of $f$ to $X$.

12. The non-Archimedean case

Instead of considering normed spaces over $\mathbb{R}$ or $\mathbb{C}$, one can also consider a normed space $X$ over a field $F$ with an absolute value. There is special interest in the case when the norm and absolute value are non-Archimedean in the sense that they each satisfy the strong or ultrametric triangle inequality

$$\|x + y\| \leq \max (\|x\|, \|y\|) \quad \text{for all } x, y \in X. \quad (1)$$

As a consequence, non-Archimedean geometry has the following properties:
(a) every point in a circle $\{y \in X: \|y - x\| \leq r\}, \ r > 0$, is a center;
(b) all 'triangles' (triples of points, that is) are isosceles; and
(c) if two circles meet, they are concentric; furthermore, any mutually intersecting collection of closed balls is totally ordered.

Non-Archimedean functional analysis gives us a chance to consider the what-if question of what would functional analysis be like without the Hahn–Banach theorem? There is a bifurcation. Non-Archimedean analysis is quite similar to ordinary analysis in situations in which the Hahn–Banach theorem holds, quite different otherwise. Nevertheless, a linear functional $f : X \to F$ is still continuous iff it is bounded on the unit ball of $X$.

Because of (c), the binary intersection property is equivalent to:

**Spherical Completeness.** Every decreasing sequence of closed balls has nonempty intersection.

$\mathbb{R}$, for example, is spherically complete. Ingleton [36] adapted Nachbin's binary intersection property characterization of extendible spaces and showed that a non-Archimedean Banach space $Y$ is extendible iff it is spherically complete. Spherical completeness is similar in appearance to completeness—namely that every decreasing sequence of closed balls whose diameters shrink to 0 has nonempty intersection—but clearly stronger. Ono [60] generalized Ingleton's result (cf. Prolla [88, p. 142]). A thorough survey of the Hahn–Banach extension property in many non-Archimedean cases is [63].

13. Ordered versions

Suppose that $X$ and $Y$ are real preordered linear spaces rather than normed spaces and that $p : X \to Y$ is a sublinear functional on $X$.

\[
\begin{align*}
\overline{A} : & \quad X & \quad \overline{A} \leq p \\
A : & \quad M \to Y & A \leq p
\end{align*}
\]
The idea now is to characterize those $Y$ for which linear extensions $\overline{A}$ dominated by $p$ on all of $X$ always exist. As shown in [37], $Y$ is ‘extendible’ in this sense iff $Y$ has the least upper bound property, that each majorized subset of $Y$ has a supremum. In the language of ordered spaces, such spaces are called (order) complete.

14. The geometric form

Planes divide $\mathbb{R}^3$ into three convex parts: the plane itself and the two ‘sides’ of the plane. Hyperplanes (see below) do a similar thing: they cleave an arbitrary real vector space into the convex subdivisions: \{x: f(x) = a\}, \{x: f(x) > a\} and \{x: f(x) < a\}. Moreover, we have (essentially) the following 1–1 correspondences:

- linear functionals $f \leftrightarrow$ Hyperplanes $H = f^{-1}(\mathbb{R})$,
- balls $B$ (open convex sets) $\leftrightarrow$ $B = U_p = \{x: p(x) < 1\}$, where $p$ is a continuous seminorm,

$H \cap U_p = \emptyset \leftrightarrow |f| \leq p.

Because of this, one can adopt a different perspective about the Hahn–Banach theorem. View it not as a statement about extendibility, but separation as in:

If a line (linear subspace) does not meet a ball (convex set) then there is a plane (hyperplane) containing the line (linear subspace) that does not meet the ball (convex set).

A form of the theorem in this guise was first proved by Mazur [7] in 1933; Bourbaki subsequently dubbed it the geometric form of the Hahn–Banach theorem.

**The Geometric Form.** In any topological vector space $X$ over $\mathbb{K}$, if the linear variety $M$ does not meet the open convex set $G$ then there is a closed hyperplane $H$ containing $M$ which does not meet $G$ either.

14.1. Separation results

Let $X'$ denote the continuous dual of the locally convex space $X$. For disjoint convex subsets $A$ and $B$ of $X$ and $f$ a real nontrivial linear form on $X$, let $H = f^{-1}(t)$ for some $t \in \mathbb{R}$. We say that $A$ and $B$ are (strictly) separated by the hyperplane $H$ if for all $a$ in $A$ and $b$ in $B$, $(f(a) < t < f(b))$ $f(a) \leq t \leq f(b)$.

(a) For distinct vectors $x$ and $y$ there exists $f \in X'$ such that $f(x) \neq f(y)$; if $x$ and $y$ are linearly independent then there exists $f$ such that $f(x) = 0$ and $f(y) = 1$.

(b) If $x$ does not meet the closed subspace $M$, then there is a continuous linear functional $f$ on $X$ which vanishes on $M$ but not on $x$.

(c) If the vector $x \notin \text{cl}\{0\}$ (topological closure of 0), then there is a continuous linear functional $f$ on $X$ such that $f(x) \neq 0$. 
(d) If $A$ and $B$ are nonempty open disjoint convex sets in the real vector space $X$ then $A$ and $B$ are strictly separated by a closed hyperplane.

(e) If $A$ and $B$ are nonempty disjoint convex subsets of $X$ with $A$ closed and $B$ compact then they are strictly separated by a closed hyperplane.

As an example of the utility of this perspective, we mention the result of James (see Holmes [34, p. 161]):

A real Banach space is reflexive if and only if each pair of disjoint closed convex subsets, one of which is bounded, can be strictly separated by a hyperplane.

15. Concluding remarks

The Hahn–Banach family of theorems more aptly describes what exists today, and it is a thriving mathematical enterprise. To name just a very few recent developments, we have:

- Ding [14]. Some conditions for a nonlocally convex space such as the $\ell_p$, $0 < p < 1$, to have the Hahn–Banach extension property.
- Plewnia [68]. Instead of a linear subspace of a real linear space $X$ let $C$ be a nonempty convex subset of $X$. Let $p : X \to \mathbb{R}$ be a convex function and let $f : C \to \mathbb{R}$ be a concave function with $f(x) \leq p(x)$ on $C$. Then there exists a linear function $g : X \to \mathbb{R}$ and a real constant $a$ such that $g(x) + a \leq p(x)$ for $x \in X$ and $f(x) \leq p(x) + a$ for $x \in C$.
- Ruan [69]. A Hahn–Banach theorem for bisublinear functionals.
- Su [83]. A Hahn–Banach theorem for a class of linear functionals on probabilistic normed spaces.

Will it ever end? The wonder is that we don’t know.

References