Lecture 5. Hahn–Banach extension theorem

Paul Delatte
delatte@usc.edu
University of Southern California

Last updated: 31 December, 2022

Theorem 1 (**Hahn–Banach Extension Theorem**). *Let X be a vector space over $\mathbb{R}$, $q\colon X \to \mathbb{R}$ a convex function on X, and E a vector subspace of X. If $f\colon E \to \mathbb{R}$ is a linear functional on E such that $f(x) \leq q(x)$ for all $x \in E$, then f can be extended to a linear functional on X satisfying $f(x) \leq q(x)$ for all $x \in X$.*

Proof. We first extend f in one dimension. Take $x \in X \setminus E$ and define $V = \operatorname{span}(x, E) = E \oplus x\mathbb{R}$. For each $v \in V$, there is a unique decomposition $v = w + \beta x$ for some $\beta \in \mathbb{R}$. If F is a linear extension of f on V, it must satisfy $F(w + \beta x) = f(w) + \beta F(x)$. We only need to choose $F(x)$ so that $F(w + \beta x) \leq q(w + \beta x)$. This is equivalent to

$$\sup_{\beta>0, w\in E} \frac{q(w - \beta x) - f(w)}{-\beta} \leq F(x) \leq \inf_{\beta>0, w\in E} \frac{q(w + \beta x) - f(w)}{\beta}.$$

We thus need to show that

$$\beta_2 f(w_1) + \beta_1 f(w_2) \leq \beta_2 q(w_1 - \beta_1 x) + \beta_1 q(w_2 + \beta_2 x)$$

for all $\beta_1, \beta_2 > 0$ and all $w_1, w_2 \in E$. If we define $\lambda = \beta_2/(\beta_1 + \beta_2)$, then

$$\begin{aligned}
\beta_2 f(w_1) + \beta_1 f(w_2) &= (\beta_1 + \beta_2) f(\lambda w_1 + (1 - \lambda) w_2) \\
&\leq (\beta_1 + \beta_2) q(\lambda w_1 + (1 - \lambda) w_2) \\
&= (\beta_1 + \beta_2) q(\lambda(w_1 - \beta_1 x) + (1 - \lambda)(w_2 + \beta_2 x)) \\
&\leq \beta_2 q(w_1 - \beta_1 x) + \beta_1 p(w_2 + \beta_2 x),
\end{aligned}$$

where the first inequality follows from domination and the second from convexity. This concludes the extension in one dimension. To conclude the proof, we invoke Zorn's lemma and use the one dimension extension to contradict maximality. Let $\mathscr{F}$ the collection of all extensions F of f satisfying $F(x) \leq q(x)$ on its domain. Then $\mathscr{F}$ can be partially ordered by inclusion with respect to the domain and every of its chains has an upper bound defined on the union of all domains. By Zorn's lemma, there is a maximal extension F^{max}. If F^{max} is not defined on X, then it can be extended to one more dimension as in the first part of the proof, which contradicts the maximality of F^{max}. $\square$

Reference. See T.4.13. in Teschl TRFA p.107 or T.3.2. in Rudin FA p.57 or T.6.3.3. in Bogachev&Smolyanov RFA p.197. The theorem emerged progressively from results of Helly (1912), Riesz (1923), Hahn (1927), and Banach (1929). Our proof is standard: it is based on transfinite induction for which we need the axiom of choice in the form of

Zorn's lemma. The proof is thus nonconstructive, and it must be the case for the axiom of choice is necessary for the Hahn–Banach theorem to hold (see R.1.5.9. in Tao RA p.69). Note that the theorem is often stated under the stronger condition that p is a sublinear functional (recall that every sublinear functional is convex).

Remark. The general purpose of the Hahn-Banach extension theorem is to guarantee that there exist sufficiently many linear functionals. The substance of the theorem is not simply to ensure that a linear extension exists (it is actually easy to do so), but to ensure that a linear extension which remains dominated by q exists. This is often sufficient to justify duality theory (i.e., the study of continuous linear functionals as a way to understanding the space of origin), notably in the context of locally convex spaces. Indeed, for the application of the Hahn–Banach extension theorem, one needs some sublinear bounds on f, but these may not be easily found; in the case of locally convex spaces, all continuous linear functionals are naturally bounded by some seminorms, hence the extension result applies readily to all continuous linear functionals.

Lemma 2 (**Zorn's Lemma**). *If X is a nonempty partially ordered set and every totally ordered subset of X has an upper bound, then X has a maximal element.*

Proof. We assume the axiom of choice and prove Zorn's lemma by contradiction. Suppose X has no maximal element. By the axiom of choice, for each totally ordered subset C of X, there is $g(C) \in X \setminus C$ such that $g(C) > x$ for all $x \in C$. We call a totally ordered subset C of X a conforming chain if it is well ordered and if for every segment $C_x = \{y \in C : y < x\}$ we have $g(C_x) = x$. We consider C as a segment of itself. If C is a nonempty conforming chain, then it has a least element $x \in C$ by well ordering and $C_x = \emptyset$, so that $g(\emptyset) \in C$. Hence, two nonempty conforming chains of X must meet. More precisely, by recursion, any nonempty conforming chain of X must start like $g(\emptyset) < g(g(\emptyset)) < \dots$. This suggests that for any two conforming chains, one must be a segment of the other. We prove this claim formally. Suppose that D is not a segment of C. We prove that $C = D_z$ for some $z \in D$. Since the chains are conforming, we have $D \setminus C \neq \emptyset$. Let $z = \min(D \setminus C)$. Then $D_z \subseteq C$. We prove that $C \subseteq D_z$. By contradiction, suppose that $C \setminus D_z \neq \emptyset$. Let $x = \min(C \setminus D_z)$ and $w = \min(D \setminus C_x)$. We prove that $D_w = C_x$. If $d \in D_w$, then $d \in D$ and $d < w$, which means by definition of w that $d \notin D \setminus C_x$, hence $d \in C_x$. If $c \in C_x$, then $c \in C$ and $c < x$, which means by definition of x that $c \notin C \setminus D_z$, hence $c \in D_z$; moreover, $c \neq w$ because $w \notin C_x$; finally, it is not possible that $c > w$, for it would imply $w \in C_x$ due to the fact that given any $a \in C_x$ and any $d \in D$ such that $a > d$, it holds that $d \in C_x$; hence $c < w$, and so $c \in D_w$. Therefore, $D_w = C_x$. Since $w = g(D_w) = g(C_x) = x$ and $x \in C$, we have $w \in C$ and thus $w \neq z$ by definition of z. Since $w \leq z$, it follows that $w < z$. Hence $x = w \in D_z$, contradicting the choice of x. Therefore $C \subseteq D_z$. We now take the union over all conforming chains of X to form C^{max} which is a conforming chain from the previous proved fact. We can add $g(C^{max}) \notin C^{max}$ to C^{max} to form another conforming chain $C^{max} \cup \{g(C^{max})\}$. Then $g(C^{max}) \in C^{max}$ by definition of C^{max}, hence a contradiction. ◻

Reference. The proof is standard: it is an adaption to Zorn's lemma of Zermelo's proof (1904) of the well ordering theorem, originally (?) by Kneser (1950). It can be found in Lewin (1991) or T.A.2. in Teschl TRFA p.514. Other proofs exist, in particular by way of ordinals.

Remark. Zorn's lemma (or the Kuratowski–Zorn lemma) is actually equivalent to the axiom of choice (the reverse implication is easy). The axiom of choice states that: the Cartesian product of nonempty sets is nonempty. Starting from the axiom of choice, Zorn's lemma was proved in 1922 by Kuratowski and independently by Zorn in 1934. Another earlier equivalent formulation of Zorn's lemma is the maximum Hausdorff principle proved in 1914 by Hausdorff which states that: every nonempty partially ordered set contains a maximal totally ordered set. Another important result equivalent to the axiom of choice is Zermelo's well-ordering theorem which states that: every nonempty set can be well ordered. The axiom of choice is an axiom of set theory independent of the ZF set theory. It is now accepted and used by most mathematicians with ZF to form ZFC as the foundation for set theory. While it is standard to use the axiom of choice without acknowledgment, it is not the case for equivalent formulations such as Zorn's lemma which is almost always explicitly mentioned when used. Along the same line, it is standard in texts to deduce Zorn's lemma from the axiom of choice, but (aware of the equivalence) the choice of the starting axiom is a matter of taste (and so, we could have stated Zorn's lemma without proof.)

Addendum. "[I]n spite (or perhaps because) of its nonconstructive proof" (Lax), the Hahn–Banach extension theorem has very concrete geometric implications: in particular, it provides conditions for the strong separation of convex sets via a hyperplane.

Lemma 3. *Let X be a vector space. For any subset $S \subseteq X$, define the Minkowski functional (or gauge) $p_S\colon X \to [0,+\infty]$ of S as $p_S(x) = \inf\{t > 0 : x \in tS\}$. If $U \subseteq X$ is a convex set containing the origin 0, then:*

1. $p_S(\lambda x) = \lambda p_S(x)$ for all $\lambda \geq 0$;

2. $p_S(x + y) \leq p_S(x) + p_S(y)$;

3. $\{x \in X : p_S(x) < 1\} \subseteq U \subseteq \{x \in X : p_S(x) \leq 1\}$;

4. if, moreover, X is a topological vector space and U is open, then $U = \{x \in X : p_S(x) < 1\}$.

Proof. 1. Let $\lambda \geq 0$. If $\lambda = 0$, the result follows immediately. Suppose $\lambda > 0$. Then $p_S(\lambda x) = \inf\{t > 0 : \lambda x \in tU\} = \inf\{t > 0 : x \in \lambda^{-1}tU\} = \lambda \inf\{\lambda^{-1}t > 0 : x \in \lambda^{-1}tU\} = \lambda p_U(x)$.

2. Let $t, s > 0$ and $x \in tU$ and $y \in sU$. Then

$$\frac{t}{t+s}\frac{x}{t} + \frac{s}{t+s}\frac{y}{s} = \frac{x+y}{t+s}$$

belongs to U by convexity. Therefore $p_U(x + y) \leq t + s$ by definition of p_U and homogeneity, and taking the infimum over all t and all s yields subadditivity.

3. Let $x \in X$ such that $p_U(x) < 1$. Then $t^{-1}x \in U$ for some $t < 1$ and thus $x \in U$ by convexity (and $0 \in U$). Similarly, let $x \in U$. Then $t^{-1}x \in U$ for some $t \geq 1$ by convexity (and $0 \in U$) and thus $p_U(x) \leq 1$.

4. If U is open and $x \in U$, then there is $\varepsilon > 0$ such that $(1 + \varepsilon)x \in U$ and thus $p_U(x) \leq (1 + \varepsilon)^{-1}$. □

Corollary 4 (Hahn–Banach Separation Theorem in TVS). *Let X be a topological vector space over $\mathbb{R}$. Let A and B be subsets of X. If A and B are disjoint nonempty convex and if A is open, then there exists a continuous linear functional $f\colon X \to \mathbb{R}$ and $\alpha \in \mathbb{R}$ such that $f(x) < \alpha \leq f(y)$ for all $x \in A$ and all $y \in B$. If B is also open, then the second inequality is strict.*

Proof. Take $a \in A$ and $b \in B$. Then $C = (A - a) - (B - b) = \{(x - a) - (y - b) : x \in A, y \in B\}$ is open (since A is open) and convex (since A and B are convex) and it contains the origin 0. By the previous lemma, the Minkowski functional of W is sublinear and thus convex. Since $A \cap B = \emptyset$, we have $c = b - a \notin C$, and so $p_C(c) \geq 1$ by the previous lemma. Define the functional $f\colon \operatorname{span}\{c\} \to \mathbb{R}$ by $f(x) = \lambda$ where $x = \lambda c$ for some unique $\lambda \in \mathbb{R}$. The functional is clearly linear. If $\lambda \geq 0$, then $f(\lambda c) = \lambda \leq \lambda p_C(c) = p_C(\lambda c)$. If $\lambda < 0$, then $f(\lambda c) < 0 \leq p_U(\lambda c)$. Thus $f \leq p$ on $\operatorname{span}\{c\}$. The conditions of the Hahn–Banach theorem are satisfied, and so f extends to a linear function on X such that $f(x) \leq p(x)$ for all $x \in X$. By linearity and domination, $|f(x)| \leq 1$ on $C \cap (-C)$. That is, f is bounded in a neighborhood of 0, and so f is continuous at 0. By translation, f is continuous on X. For any $z \in C$, we have $p_C(z) < 1$, hence $f(x) - f(y) + 1 = f(x - y + c) \leq p_C(x - y + c) < 1$ for all $x \in A$ and all $y \in B$, and so $f(x) < f(y)$ for all $x \in A$ and all $y \in B$. Therefore $f(A)$ and $f(B)$ are disjoint convex subsets of $\mathbb{R}$. Suppose now that there is some $a_0 \in A$ such that $f(a_0) = \sup f(A)$. Then by continuity of $\lambda \mapsto a_0 + \lambda c$, there is some $\varepsilon > 0$ such that $a_0 + \varepsilon c \in U$. Then $f(a_0) + \varepsilon = f(a_0 + \varepsilon c) \leq f(a_0)$, a contradiction. Hence, for $\alpha = \sup f(A)$, we have $f(x) < \alpha \leq f(y)$ for all $x \in A$ and all $y \in B$. If B is also open, a similar argument shows that $\inf f(B) < f(y)$ for all $y \in B$. □

Corollary 5 (Hahn–Banach Separation Theorem in LCTVS). *Let X be a locally convex topological vector space over $\mathbb{R}$. Let A and B be subsets of X. If A and B are disjoint nonempty convex and if A is compact and B is closed, then then there exists a continuous linear functional $f\colon X \to \mathbb{R}$ and $\alpha \in \mathbb{R}$ such that $f(x) < \alpha < f(y)$ for all $x \in A$ and all $y \in B$.*

Proof. If we prove that there exists a convex neighborhood U of 0 such that $(A + U) \cap (B + U) = \emptyset$, then we can apply the Hahn–Banach separation theorem in TVS for $A + U$ and $B + U$ and get the result. If we find a convex balanced neighborhood V of 0 such that $(A + V) \cap B = \emptyset$, then we can take $U = V/2$. Since B is closed and X is locally convex, for every $x \in X$, there exists a convex balanced neighborhood V_x of 0 such that $(x + V_x) \cap B = \emptyset$. The family of sets $\{x + 4^{-1}V_x : x \in A\}$ forms an open cover of A, of which we extract a finite subcover $x_1 + 4^{-1}V_{x_1}, \ldots, x_n + 4^{-1}V_{x_n}$ by compactness. Then we can take $V = \bigcap_{i=1}^n 4^{-1}V_{x_i}$. This concludes the proof. □

Reference. The proof of the properties of the Minkwoski functional follows L.5.1. in Teschl TRFA p.134 or T.1.35. in Rudin FA p.26. The proofs of the two geometric versions theorem are standard. See T.3.4. in Rudin FA p.59 or T.5.2 and C.5.4. in Teschl TRFA p.134-136 or T.6.3.7-8. and C.8.3.5. in Bogachev&Smolyanov p.200-201 and p.367.

References

BOGACHEV, V. I., AND O. G. SMOLYANOV (2020): *Real and functional analysis*. Springer.

KNESER, H. (1950): "Eine direkte Ableitung des Zirnschen lemmas aus dem Auswahlaxiom," *Mathematische Zeitschrift*, 53(2), 110–113.

LAX, P. D. (2002): *Functional analysis*, vol. 55. John Wiley & Sons.

LEWIN, J. W. (1991): "A simple proof of Zorn's lemma," *The American mathematical monthly*, 98(4), 353.

RUDIN, W. (1973): *Functional analysis*. McGraw-Hill.

TESCHL, G. (2019): "Topics in real and functional analysis," *Unpublished lecture notes*.