A more intuitive proof of Hahn–Banach

Consider the real vector space version. We want to extend λ to $\tilde\lambda$ on $\tilde Y := \operatorname{span}(Y, z)$, where $z \notin Y$ is fixed. We want to preserve the inequality $\tilde\lambda(x) \le p(x)$. Since any element $\tilde y \in \tilde Y$ has the form $\tilde y = y + \beta z$, and

$$\tilde\lambda(\tilde y) = \lambda(y) + \beta\tilde\lambda(z),$$

we just need to specify $\tilde\lambda(z)$.

The inequality $\tilde\lambda(\tilde y) \le p(\tilde y)$ is equivalent to

$$\lambda(y) + \beta\tilde\lambda(z) \le p(y + \beta z). \tag{1}$$

Solve for $\tilde\lambda(z)$:

$$\tilde\lambda(z) \le \frac{1}{\beta}\big(p(y+\beta z) - \lambda(y)\big) \quad \text{when } \beta > 0,$$

and

$$\frac{1}{\beta}\big(p(y+\beta z) - \lambda(y)\big) \le \tilde\lambda(z) \quad \text{when } \beta < 0.$$

To get closer to the notation in the book, set $-\beta = \alpha > 0$. Then we get the requirement

$$\frac{1}{\alpha}\big(-p(y-\alpha z) - \lambda(y)\big) \le \tilde\lambda(z) \le \frac{1}{\beta}\big(p(y+\beta z) - \lambda(y)\big), \quad \forall\alpha > 0,\ \forall\beta > 0,\ \forall y \in Y. \tag{2}$$

Now, varying α, β, z gives us a family of intervals where $\tilde\lambda(z)$ must belong but what if this family has an empty intersection? The chain of inequalities in the book shows that they do not. More precisely, we treat y on the left and on the right as two different y's, calling them y_1 and y_2, and as in the book, show that

$$\frac{1}{\alpha}\big(-p(y_1+\alpha z) + \lambda(y_1)\big) \le \frac{1}{\beta}\big(p(y_2+\beta z) - \lambda(y_2)\big), \quad \forall\alpha > 0,\ \forall\beta > 0,\ \forall y_1, y_2 \in Y.$$

Note that this uses the convexity of p. Then we fix (α, y_1) on the left, and take the infimum on the right over all $(\beta, y_2) \in \mathbb{R}_+ \times Y$. Then we take the supremum on the left over all $(\alpha, y_1) \in \mathbb{R}_+ \times Y$. This gives us the inequality on the top of p. 76 which is inequality between two numbers. This shows that the intersection of all those intervals is non-empty. Therefore, we can insert $\tilde\lambda(z)$ inside. This would satisfy the requirement (2), and therefore, (1), because all conditions so far were if and only if.

Remarks:

Remark 1. When X is Hilbert, the theorem is trivial when $p(x) = A|x|$, $A > 0$. Just write $x = y + z$, $y \in Y$, $z \in Y^\perp$, and declare $\Lambda = 0$ on $Y^\perp$, i.e., $\Lambda x = \lambda y$. Then $|\Lambda x| = |\lambda y| \le A\|y\| \le A\|x\|$. In fact, in this case the extension is unique.

Remark 2. When X is real, and $p(x) = A|x|$, $A > 0$, you may wonder why don't we have $|\lambda(x)| \le p(x)$ (with absolute value), which is equivalent to λ being bounded. In fact, applying $\lambda(x) \le p(x)$ to x and $-x$ yields that. The complex version of the theorem is formulated with absolute values anyway.

Remark 3. About Corollary 2: Assume for simplicity that we want $\|\Lambda\|_{X^*} = 1$. We can always do that by dividing Λ by its norm. In fact, the proof provides Λ with norm one. Then we want to prove that there is a bounded linear functional with the property $\Lambda y = \|y\|$ for that special $y \ne 0$ (the case $y = 0$ is trivial). We start with $\lambda(ay) = a\|y\|$ which clearly has norm one on the span of y by definition of a norm. Then we extend it to X.

What if X is Hilbert? Then Λx would be just the orthogonal projection of x to (the span of) y, given by $\Lambda x = (y/\|y\|, x)$ when $y \neq 0$. In fact, the whole point of this corollary is to show that such scalar projections can still be defined on Banach spaces, not necessarily in an unique way.

Remark 4. About Corollary 3: Let us start with the Hilbert case again. When $\operatorname{dist}(y, Z) = d$, there is a closest element $z \in Z$ to y, realizing that distance, by the geometry of the Hilbert spaces. Look at Figure II.1 in the book, where $M = Z$, $z = y$, and $|w| = d$. Then Λy would be the orthogonal scalar projection to $w = y - z \perp Z$, normalized. Clearly, it would have the required properties. The whole point of this corollary is to prove that such a scalar projection exists in Banach spaces but we are not claiming now that $\operatorname{dist}(y, Z) = \inf_{z \in Z} \|y - z\|$ is attained for some z, and of course, there is no orthogonality. Instead of that, one can study the quotient X/Z, playing the role of $Z^\perp$, etc.

Remark 5. The example I mentioned in class: $C([-1, 1]) \subset L^\infty([-1, 1])$, both with the sup norm. We can extend $\lambda(f) = f(0)$ from $C([-1, 1])$ to $L^\infty([-1, 1])$ with the same norm. Let us first extend it to the span of $C([-1, 1])$ and $\{h\}$, where h is the Heaviside function. This gives us all piece-wise continuous functions with a jump discontinuity at 0 only. We can take

$$\tilde{\lambda}(f) = af(0-) + bf(0+),$$

where $a \geq 0$, $b \geq 0$, $a + b = 1$ are fixed. One can see directly that this does the job. Following the proof of Hahn-Banach, we can see that those are all choices. It is a good exercise to see how this follows from (2).

Next, there is nothing much special about the choice of h (Heaviside) there. One could choose any piece-wise continuous function with a non-zero jump at 0.

Finally, consider the distance d from h to $C([-1, 1])$. It is easy to see that $d = 1/2$. What are the closest continuous function to h (staying at distance one), and do they actually exist? One of them is $f = 1/2$. But there are many more. We need $f(0) = 1/2$ and $\|f - h\| = 1/2$. A small enough perturbation keeping the value $1/2$ at 0 would do it.