Version of 14.9.18

Kirszbraun's theorem

D.H.Fremlin

University of Essex, Colchester, England

Wikipedia gives a statement of this theorem and an outline of its history, but no online source for the proof. It's so pretty that I write one out here.

1 The essential ideas

1A The context I'll come to the actual statement of the theorem, taken from `http://en.wikipedia.org/wiki/Kirszbraun_theorem`, in 1G below. This will be in the standard full-generality form concerning Lipschitz maps between (real) Hilbert spaces H_1 and H_2. If you aren't familiar with 'Hilbert spaces' (`http://en.wikipedia.org/wiki/Hilbert_space`), then you will probably prefer to start by taking both H_1 and H_2 to be a Euclidean space $\mathbb{R}^n$, with the inner product

$$(x|y) = x\,.\,y = \sum_{i=1}^{n} \xi_i \eta_i$$

if $x = (\xi_1, \ldots, \xi_n)$ and $y = (\eta_1, \ldots, \eta_n)$ belong to $\mathbb{R}^n$, and the norm

$$\|x\| = \sqrt{(x|x)} = \sqrt{\sum_{i=1}^{n} \xi_i^2},$$

so that

$$\|x - y\| = \sqrt{\sum_{i=1}^{n} (\xi_i - \eta_i)^2}$$

is the Euclidean distance from x to y calculated with the n-dimensional version of Pythagoras' theorem. In fact all the really interesting ideas of the proof, in 1B-1F below, are needed for the two-dimensional case $n = 2$. (The case $n = 1$ is much easier, as noted in Wikipedia.) I do have to warn you, however, that ordinary proofs of Kirszbraun's theorem involve 'Tychonoff's theorem' (see 2B below), and unless the words 'topology', 'compact set' and 'continuous function' mean something to you, you are going to have a good deal of work to do to understand the 'first proof' offered in §2. There is an alternative method using 'filters' which I will describe in §3, but this also demands some highly abstract reasoning on top of basic real analysis.

1B The miraculous bit: Lemma Let H_1 and H_2 be Hilbert spaces. Suppose that J is a non-empty finite subset of H_1, and $g : J \to H_2$ a function such that $\|g(x) - g(y)\| \le \|x - y\|$ and $\|g(x)\| > \|x\|$ for all x, $y \in J$. Then 0 does not belong to the convex hull $\Gamma(g[J])$ of $g[J] = \{g(x) : x \in J\}$.

proof Note first that, for any x, $y \in J$,

$$\begin{aligned}(x|y) &= \frac{1}{2}(\|x\|^2 + \|y\|^2 - \|x - y\|^2)\\ &< \frac{1}{2}(\|g(x)\|^2 + \|g(y)\|^2 - \|g(x) - g(y)\|^2) = (g(x)|g(y))\end{aligned}$$

because $\|x\| < \|g(x)\|$, $\|y\| < \|g(y)\|$ and $\|x - y\| \ge \|g(x) - g(y)\|$. Now suppose that $w \in \Gamma(g[J])$. Then there is a family $\langle \lambda_x \rangle_{x \in J}$ of non-negative real numbers such that $\sum_{x \in J} \lambda_x = 1$ and $w = \sum_{x \in J} \lambda_x g(x)$. So

$$\begin{aligned}\|w\|^2 = (w|w) &= (\sum_{x \in J} \lambda_x g(x) | \sum_{y \in J} \lambda_y g(y)) = \sum_{x,y \in J} \lambda_x \lambda_y (g(x)|g(y))\\ &> \sum_{x,y \in J} \lambda_x \lambda_y (x|y)\end{aligned}$$

(because $(g(x)|g(y)) > (x|y)$, therefore $\lambda_x \lambda_y (g(x)|g(y)) \ge \lambda_x \lambda_y (x|y)$, for all x, $y \in J$, and there is at least one $x \in J$ such that $\lambda_x > 0$, so that $\lambda_x \lambda_x (g(x)|g(x)) > \lambda_x \lambda_x (x|x)$)

$$= (\sum_{x \in J} \lambda_x x | \sum_{y \in J} \lambda_y y) = \| \sum_{x \in J} \lambda_x x \|^2 \geq 0,$$

and $w \neq 0$.

1C Basic facts about Hilbert spaces: Lemma Let H be a Hilbert space.

(a)(i) For any x, $y \in H$, $|(x|y)| \leq \|x\|\|y\|$.

(ii) For any $x \in H$, $\|x\| = \max\{(x|c) : c \in H, \|c\| \leq 1\}$.

(b) For any x, $y \in H$, $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$.

(c) If $C \subseteq H$ is a non-empty closed convex set, and $b \in H$, then there is a $b' \in C$ such that $(z-b|b'-b) \geq \|b'-b\|^2$ for every $z \in C$.

(d) If $f : H \to \mathbb{R}$ is a linear functional and $\gamma = \sup\{|f(x)| : x \in H, \|x\| \leq 1\}$ is finite, there is a unique $c \in H$ such that $f(x) = (x|c)$ for every $x \in H$.

(e) If $I \subseteq H$ is finite, then $\Gamma(I)$ is compact for the norm topology on H.

Remark As I said in 1A, if you don't know what a Hilbert space is, just take H to be $\mathbb{R}^n$ for some n – in fact, the case $n = 2$ is already enough to use every idea here.

proof (a)(i) (This is the 'Cauchy-Schwarz inequality'.) If either x or y is zero then we have 0 on both sides. Otherwise, set $\alpha = \|x\|$ and $\beta = \|y\|$ and consider

$$\begin{aligned} 0 &\leq \|\beta x - \alpha y\|^2 = (\beta x - \alpha y|\beta x - \alpha y) \\ &= \beta^2\|x\|^2 - 2\alpha\beta(x|y) + \alpha^2\|y\|^2 = 2\alpha^2\beta^2 - 2\alpha\beta(x|y); \end{aligned}$$

dividing by $2\alpha\beta > 0$, $0 \leq \alpha\beta - (x|y)$ and $(x|y) \leq \alpha\beta = \|x\|\|y\|$. Similarly,

$$-(x|y) = (-x|y) \leq \|-x\|\|y\| = \|x\|\|y\|,$$

so $|(x|y)| \leq \|x\|\|y\|$.

(ii) In particular, if $\|c\| \leq 1$, then $(x|c) \leq \|x\|$. In the other direction, if $x = 0$ then $\|x\| = 0 = (x|0)$, while otherwise we can set $c = \frac{1}{\|x\|}x$ and see that $\|c\| = 1$ and $(x|c) = \frac{1}{\|x\|}\|x\|^2 = \|x\|$.

(b)

$$\begin{aligned} \|x+y\|^2 + \|x-y\|^2 &= (x+y|x+y) + (x-y|x-y) \\ &= \|x\|^2 + 2(x|y) + \|y\|^2 + \|x\|^2 - 2(x|y) + \|y\|^2 = 2\|x\|^2 + 2\|y\|^2. \end{aligned}$$

(c)(i) The set $\{\|z-b\| : z \in C\} \subseteq \mathbb{R}$ is non-empty and has a lower bound, so it has a greatest lower bound α say. For each $n \in \mathbb{N}$, set $C_n = \{z : z \in C, \|z-b\| \leq \alpha + 4^{-n}\}$. Then $\|z - z'\| \leq 2^{-n}\sqrt{8\alpha+4}$ for all z, $z' \in C_n$. To see this, note that because C is convex, $w = \frac{1}{2}(z+z')$ belongs to C, so $\|w-b\| \geq \alpha$. Now

$$\begin{aligned} 4\alpha^2 &= \|2w - 2b\|^2 = \|(z_1 - b) + (z_2 - b)\|^2 \\ &= 2\|z_1 - b\|^2 + 2\|z_2 - b\|^2 - \|(z_1 - b) - (z_2 - b)\|^2 \end{aligned}$$

(by (b) just above)

$$\leq 4(\alpha + 4^{-n})^2 - \|z_1 - z_2\|^2,$$

so

$$\|z_1 - z_2\|^2 \leq 4((\alpha + 4^{-n})^2 - \alpha^2) = 8\alpha \cdot 4^{-n} + 4 \cdot 4^{-2n} \leq 4^{-n}(8\alpha + 4)$$

and $\|z_1 - z_2\| \leq 2^{-n}\sqrt{8\alpha+4}$.

(ii) Since every C_n is non-empty, we can choose a sequence $\langle z_n \rangle_{n \in \mathbb{N}}$ in H such that $z_n \in C_n$ for every $n \in \mathbb{N}$. In this case, $\|z_m - z_n\| \leq 2^{-m}\sqrt{8\alpha+4}$ whenever $m \leq n$, because in this case both z_m and z_n belong to C_m. So $\langle z_n \rangle_{n \in \mathbb{N}}$ is a Cauchy sequence and has a limit b' in H.[1] Because every z_n belongs to C, and C is closed, $b' \in C$; and now

[1] This is where we need to know that H is a Hilbert space, rather than just an inner product space.

$$\alpha \le \|b'-b\| = \lim_{n\to\infty} \|z_n - b\| \le \alpha,$$

so $\|b'-b\| = \alpha$.

(iii) Next, for $z \in C$, we know that $\delta z + (1-\delta)b' \in C$ for every $\delta \in]0,1]$, because C is convex. But this means that

$$\begin{aligned}\|b'-b\|^2 &\le \|\delta z + (1-\delta)b' - b\|^2 = \|\delta(z-b) + (1-\delta)(b'-b)\|^2 \\ &= \delta^2\|z-b\|^2 + 2\delta(1-\delta)(z-b|b'-b) + (1-\delta)^2\|b'-b\|^2;\end{aligned}$$

subtracting $\|b'-b\|^2$ from both sides,

$$0 \le \delta^2\|z-b\|^2 + 2\delta(1-\delta)(z-b|b'-b) - 2\delta\|b'-b\|^2 + \delta^2\|b'-b\|^2;$$

dividing by $2\delta > 0$,

$$0 \le \frac{1}{2}\delta\|z-b\|^2 + (1-\delta)(z-b|b'-b) - \|b'-b\|^2 + \frac{1}{2}\delta\|b'-b\|^2;$$

rearranging the terms,

$$\|b'-b\|^2 - (z-b|b'-b) \le \frac{1}{2}\delta(\frac{1}{2}\|z-b\|^2 - 2(z-b|b'-b) + \|b'-b\|^2);$$

letting $\delta \downarrow 0$,

$$\|b'-b\|^2 - (z-b|b'-b) \le 0,$$

that is,

$$(z-b|b'-b) \ge \|b'-b\|^2.$$

(d)(i) I'll start by checking that there can be at most one c with this property. If $f(x) = (x|c) = (x|c')$ for every $x \in H$, then

$$\|c-c'\|^2 = (c-c'|c-c') = (c-c'|c) - (c-c'|c') = f(c-c') - f(c-c') = 0,$$

and $c - c' = 0$, that is, $c = c'$.

So I just have to show that there is some c which will serve.

(ii) If $f(x) = 0$ for every $x \in H$, we can, and must, take $c = 0$. So from now on I will suppose that we have a $b \in H$ such that $f(b) \ne 0$.

(iii) Set $C = \{x : f(x) = 0\}$. Then C is a linear subspace of H (because $f(0) = 0$, so $0 \in C$, and if x, $y \in C$ and α, $\beta \in \mathbb{R}$ then $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) = 0$, so $\alpha x + \beta y \in C$). In particular, C is a non-empty convex set.

It will help to know that $|f(x)| \le \gamma\|x\|$ for every $x \in H$; this is certainly true if $x = 0$, since then $|f(x)| = 0 = \gamma\|x\|$; and for non-zero x we have $\|\frac{1}{\|x\|}x\| = 1$, so that $\frac{1}{\|x\|}|f(x)| = |f(\frac{1}{\|x\|}x)| \le \gamma$ and $|f(x)| \le \gamma\|x\|$. In particular, $\gamma\|b\| \ge |f(b)| > 0$ and $\gamma > 0$.

(iv) Next, C is closed. For if x is any point of $H \setminus C$, and $\|y - x\| \le \frac{1}{2\gamma}|f(x)|$, then

$$|f(y) - f(x)| = |f(y-x)| \le \gamma\|y-x\| \le \frac{1}{2}|f(x)| < |f(x)|$$

and $f(y) \ne 0$, that is, $y \notin C$. Thus any point of $H \setminus C$ is the centre of a non-trivial ball included in $H \setminus C$, and $H \setminus C$ is open, that is, C is closed.

(v) We can therefore appeal to (c) just above to find a $b' \in C$ such that $(z-b|b'-b) \ge \|b'-b\|^2$ for every $z \in C$. In this case

$$\begin{aligned}(z-b'|b'-b) &= ((z-b)-(b'-b)|b'-b) = (z-b|b'-b) - (b'-b|b'-b) \\ &= (z-b|b'-b) - \|b'-b\|^2 \ge 0\end{aligned}$$

for every $z \in C$. In fact, if $z \in C$, then

$$(z|b'-b) = ((z+b')-b'|b'-b) \geq 0$$

because $z+b' \in C$. Consequently we also have

$$-(z|b'-b) = (-z|b'-b) \geq 0$$

for every $z \in C$, because then $-z$ also belongs to C. Thus $(z|b'-b) = 0$ for every $z \in C$. In particular, $(b'|b'-b) = 0$, so

$$(b|b'-b) = (b-b'|b'-b) = -(b'-b|b'-b) = -\|b'-b\|^2.$$

(vi) Since $f(b) \neq 0$, $b \neq b'$, $\|b'-b\| > 0$, $\beta = -\frac{f(b)}{\|b'-b\|^2}$ is defined and we can try $c = \beta(b'-b)$. In this case,

$$(b|c) = \beta(b|b'-b) = -\beta\|b'-b\|^2 = f(b),$$

while

$$(z|c) = \beta(z|b'-b) = 0$$

whenever $z \in C$.

(vii) Finally, given $x \in H$, consider $x' = x - \frac{f(x)}{f(b)}b$. We have

$$f(x') = f(x) - \frac{f(x)}{f(b)}f(b) = 0,$$

so $x' \in C$ and

$$(x|c) = (x-x'|c) + (x'|c) = (\frac{f(x)}{f(b)}b|c) + 0 = \frac{f(x)}{f(b)}(b|c) = f(x).$$

Thus we have found a suitable c.

(e) The hypercube $[0,1]^I$ is compact, and the simplex

$$S = \{\langle\lambda_x\rangle_{x\in I} : \lambda_x \geq 0 \text{ for every } x \in I, \sum_{x\in I}\lambda_x = 1\},$$

being a closed subset of $[0,1]^I$, is also compact; now the function $\langle\lambda_x\rangle_{x\in I} \mapsto \sum_{x\in I}\lambda_x x$ is a continuous surjection from S onto K, so K also is compact.

1D The key to the door: Lemma Let H_1 and H_2 be Hilbert spaces, $I \subseteq H_1$ a finite set, $f : I \to H_2$ a function such that $\|f(x)-f(y)\| \leq \|x-y\|$ for all $x, y \in I$, and a any point of H_1. Then there is a $b \in H_2$ such that $\|b-f(x)\| \leq \|a-x\|$ for every $x \in I$.

proof (a) If $I = \emptyset$ we can set $b = 0$; if $a \in I$ we can set $b = f(a)$. So let us suppose from now on that I is not empty and $a \notin I$.

(b) Set $K = \Gamma(f[I])$. Then K is a non-empty convex subset of H_2. Also K is compact for the norm topology of H_2, by Lemma 1Ca. For each $x \in I$, the functions $z \mapsto \|z-f(x)\|$ and $z \mapsto \frac{\|z-f(x)\|}{\|a-x\|}$ from K to $[0,\infty[$ are continuous. Again because I is finite, the function $h : K \to [0,\infty[$ defined by saying that $h(z) = \max_{x\in I}\frac{\|z-f(x)\|}{\|a-x\|}$ is continuous. So h attains its infimum; let $b \in K$ be such that $h(b) \leq h(z)$ for every $z \in K$. Set

$$\gamma = h(b), \quad J = \{x : x \in I, \frac{\|b-f(x)\|}{\|a-x\|} = \gamma\}.$$

Of course J is a non-empty subset of I.

(c) We find that b has to belong to $\Gamma(f[J])$. To see this, argue by contradiction: suppose that $b \notin \Gamma(f[J])$. Now $\Gamma(f[J])$ is a compact convex subset of H_2, by Lemma 1Ca again. By Lemma 1Cb, there is a $b' \in \Gamma(f[J])$ such that $(z-b|b'-b) \geq \|b'-b\|^2$ for every $z \in \Gamma(f[J])$; in particular, $(f(x)-b|b'-b) \geq \|b'-b\|^2$ for every $x \in J$.

Now consider $b_\delta = (1-\delta)b + \delta b' = b + \delta(b'-b)$ for small $\delta > 0$. We always have $b_\delta \in K$ because b and b' belong to the convex set K. If $x \in J$, then

$$(f(x)-b|b_\delta-b)=\delta f(x)-bb'-b\ge\delta\|b'-b\|^2,$$

so

$$\begin{aligned}\|f(x)-b_\delta\|^2&=\|(f(x)-b)-(b_\delta-b)\|^2\\&=\|f(x)-b\|^2-2(f(x)-b|b_\delta-b)+\|b_\delta-b\|^2\\&\le\|f(x)-b\|^2-2\delta\|b'-b\|^2+\delta^2\|b'-b\|^2<\|f(x)-b\|^2\end{aligned}$$

whenever $0<\delta\le 1$. So

$$\tfrac{\|f(x)-b_\delta\|}{\|a-x\|}<\tfrac{\|f(x)-b\|}{\|x-a\|}=\gamma$$

whenever $0<\delta\le 1$. On the other hand, for $x\in I\setminus J$, we know that

$$\lim_{\delta\downarrow 0}\tfrac{\|f(x)-b_\delta\|}{\|x-a\|}=\tfrac{\|f(x)-b\|}{\|x-a\|}<\gamma,$$

so there is a $\delta_x>0$ such that $\frac{\|f(x)-b_\delta\|}{\|x-a\|}<\gamma$ whenever $0<\delta\le\delta_x$. Because $I\setminus J$ is finite, we can now find a $\delta>0$ such that $\delta\le\delta_x$ for every $x\in I\setminus J$. But we shall now have

$$\tfrac{\|f(x)-b_\delta\|}{\|x-a\|}<\gamma$$

for every $x\in I$, so that $h(b_\delta)<\gamma=h(b)$. And we chose b to minimise h, so this ought to be impossible.

(d) Thus $b\in\Gamma(f[J])$. We can therefore apply Lemma 1B, as follows. Set $J'=\{x-a:x\in J\}$. Define $g:J'\to H_2$ by setting $g(x)=f(x+a)-b$ for $x\in J'$. Note that if x, $y\in J'$, then

$$\begin{aligned}\|g(x)-g(y)\|&=\|(f(x+a)-b)-(f(y+a)-b)\|=\|f(x+a)-f(y+a)\|\\&\le\|(x+a)-(y+a)\|=\|x-y\|.\end{aligned}$$

$$\|g(x)\|=\|f(x+a)-b\|=\|b-f(x+a)\|=\gamma\|x\|.$$

Because b belongs to $\Gamma(f[J])$, we can express it as $\sum_{x\in J}\lambda_x f(x)$ where $\lambda_x\ge 0$ for every $x\in J$ and $\sum_{x\in J}\lambda_x=1$. In this case

$$\begin{aligned}\sum_{x\in J'}\lambda_{x+a}g(x)&=\sum_{x\in J'}\lambda_{x+a}(f(x+a)-b)=\sum_{x\in J}\lambda_x(f(x)-b)\\&=\sum_{x\in J}\lambda_x f(x)-\sum_{x\in J}\lambda_x b=b-b=0,\end{aligned}$$

while of course $\lambda_{x+a}\ge 0$ for every $x\in J'$ and $\sum_{x\in J'}\lambda_{x+a}=\sum_{x\in J}\lambda_x=1$. So $0\in\Gamma(g[J'])$. By Lemma 1B, there must be an $x\in J'$ such that $\|x\|\ge\|g(x)\|=\gamma\|x\|$; that is, there is an $x\in J$ sucy that $\|x-a\|\ge\gamma\|x-a\|$. But we decided long ago, in (a) above, that we were looking only at the case in which $a\notin I$, so that $x-a\ne 0$ and $\gamma\le 1$.

(e) Finally, returning to the definition of γ in (b), we see that $h(b)\le 1$, that is, $\|f(x)-b\|\le\|x-a\|$ for every $x\in I$, so that we have the point we need.

1E Corollary Let H_1 and H_2 be Hilbert spaces, $I\subseteq H_1$ a finite set, $\gamma\in[0,\infty[$ a non-negative real number, $f:I\to H_2$ a function such that $\|f(x)-f(y)\|\le\gamma\|x-y\|$ for all x, $y\in I$, and a any point of H_1. Then there is a $b\in H_2$ such that $\|b-f(x)\|\le\gamma\|a-x\|$ for every $x\in I$.

proof If $\gamma=0$ then f must be constant and we can take b to be the constant value of f (or 0 if I is empty). Otherwise, set $g(x)=\dfrac{1}{\gamma}f(x)$ for $x\in I$. Then

$$\|g(x)-g(y)\|=\frac{1}{\gamma}\|f(x)-f(y)\|\le\|x-y\|$$

for $x \in I$. By Lemma 1E, there is a $b_1 \in H_2$ such that $\|b_1 - g(x)\| \le \|a - x\|$ for every $x \in I$. Set $b = \gamma b_1$; then $\|b - f(x)\| = \gamma\|b_1 - g(x)\| \le \gamma\|a - x\|$ for every $x \in I$.

1F Corollary Let H_1 and H_2 be Hilbert spaces, I, $J \subseteq H_1$ finite sets, $\gamma \in [0, \infty[$ and $f : I \to H_2$ a function such that $\|f(x) - f(y)\| \le \gamma\|x - y\|$ for all x, $y \in I$. Then there is a function $g : I \cup J \to H_2$ such that $g(x) = f(x)$ for every $x \in I$ and $\|g(x) - g(y)\| \le \gamma\|x - y\|$ for all x, $y \in I \cup J$.

proof Induce on the number $\#(J)$ of elements of J. For the base step, when $\#(J) = 0$, we have $J = \emptyset$ and we can take $g = f$. For the inductive step to $\#(J) = n + 1$ where $n \in \mathbb{N}$, take any $a \in J$, and set $J' = J \setminus \{a\}$. Then $\#(J') = n$, so the inductive hypothesis tells us that there is a function $f_1 : I \cup J' \to H_2$ such that $f_1(x) = f(x)$ for every $x \in I$ and $\|f_1(x) - f_1(y)\| \le \gamma\|x - y\|$ whenever x, $y \in I \cup J'$. If $a \in I \cup J'$ then $I \cup J = I \cup J'$ and we can take $g = f_1$. Otherwise, Lemma 1E, applied to f_1, tells us that there is a $b \in H_2$ such that $\|f_1(x) - b\| \le \gamma\|x - a\|$ for every $x \in I \cup J'$. So if we define $g : I \cup J \to H_2$ by setting

$$\begin{aligned} g(x) &= f_1(x) \text{ if } x \in I \cup J', \\ &= b \text{ if } x = a, \end{aligned}$$

then we shall have

$$g(x) = f_1(x) = f(x) \text{ for every } x \in I,$$

$$\begin{aligned} \|g(x) - g(y)\| &= \|f_1(x) - f_1(y)\| \le \gamma\|x - y\| \text{ if } x,\ y \in I \cup J', \\ &= \|f_1(x) - b\| \le \gamma\|x - a\| = \gamma\|x - y\| \text{ if } x \in I \cup J' \text{ and } y = a, \\ &= \|b - f_1(y)\| = \|f_1(y) - b\| \le \gamma\|y - a\| = \gamma\|x - y\| \\ &\qquad \text{if } x = a \text{ and } y \in I \cup J', \\ &= \|b - b\| = 0 = \gamma\|x - x\| \text{ if } x = a \text{ and } y = a. \end{aligned}$$

So g has the required property and the induction continues.

1G Kirszbraun's Theorem Let H_1 and H_2 be Hilbert spaces, $A \subseteq H_1$ a set and $f : A \to H_2$ a function. Suppose that $\gamma \ge 0$ is such that $\|f(x) - f(y)\| \le \gamma\|x - y\|$ for all x, $y \in A$. Then there is a function $g : H_1 \to H_2$ such that $g(x) = f(x)$ for every $x \in A$ and $\|g(x) - g(y)\| \le \gamma\|x - y\|$ for all x, $y \in H_1$.

1H Miracles and keys I'll come to actual proofs of this theorem in §§2-3 below. But I ought to explain why I've labelled the paragraphs in the way that I have. 1G is a 'theorem' because it's the target of this whole note; it's a striking, useful and non-obvious fact, which makes it a prize to grasp and hold. 1C is a list of 'basic facts' because if you have done anything with Hilbert spaces you should know most of them, and if you hope to do anything with Hilbert spaces (even the finite-dimensional ones) you should put them all in your tool-box. 1E and 1F are 'corollaries' because I think there is a chance that they will be pretty well obvious; 1E is just a re-scaling of 1D, and 1F is a natural induction. (Of course when I say 'natural', I don't mean that anyone is born with an instinct to do this sort of thing, in the way that a baby is born with instincts to grasp and suckle. I mean that after you have had a bit of training, the word 'finite set' will trigger an impulse to look in succession at the cases of sets with 0, 1 and 2 elements, and try to see if one will help with the next.)

Now between 1B and 1D I make a further distinction. When I say that 1D is the 'key', what I mean is, that an experienced pure mathematician (for whom at least one of §2 or §3 below should be essentially obvious) will see at once that it has got to be true (if the target theorem is true at all) and that with this established, the rest ought to be mopping up. Furthermore, the idea in part (a) of the proof, picking b to minimize $\max_{x \in I} \frac{\|b - f(x)\|}{\|a - x\|}$, may not come instantly to mind, but is easy to remember and is the kind of trick which often works. But even with these hints, the proof of Lemma 1D can present real difficulties if you do not have a friend or a book to point you to a version of the idea in 1B. And 1B doesn't remind me of anything I have seen anywhere else. That is why I call it a 'miracle'. Of course it is no more miraculous than Pythagoras' theorem. But it does stand by itself; it is a fact about the geometry of Euclidean space for which I do not have a picture to show me why it works.

2 First proof

In 1H, I said that with Lemma 1D and its corollary 1F in hand the rest of the proof of Kirszbraun's theorem is mopping up. Essentially there is just one way of doing this, but it can be expressed in more than one way. If you have done a conventional first course in functional analysis, with Tychonoff's theorem and the Banach-Alaoglu theorem, the following is likely to be the most natural approach. If not, you may find that the technique in §3 below is more directly accessible,

2A Weak topologies Let H be a Hilbert space. Then H has a **weak topology** for which all the functionals $x \mapsto (x|y)$, for $y \in H$, are continuous, and every ball $B = \{x : \|x\| \le \alpha\}$, for $\alpha \ge 0$, is compact.

Remark This is really a special case of the Banach-Alaoglu theorem, because if we identify H with its own dual space (using Lemma 1Cd) then its weak topology is just its weak* topology. If this remark doesn't make much sense to you, then note that if $H = \mathbb{R}^n$, as considered in 1A above, then you just have to know that the functionals $x \mapsto (x|y) = x \centerdot y$ are continuous in the usual sense, and the balls $\{x : \|x - b\| \le \alpha\}$ are compact in the sense of the n-dimensional Heine-Borel and Bolzano-Weierstrass theorems.

2B Tychonoff's theorem (`http://en.wikipedia.org/wiki/Tychonoff's_theorem`) Suppose that we are given any family $\langle X_i \rangle_{i \in I}$ of compact spaces. Let $X = \prod_{i \in I} X_i$ be the set of functions g defined on I such that $g(i) \in X_i$ for every $i \in I$. Then X has a **product topology** for which all the functions $g \mapsto g(i) : X \to X_i$ are continuous and X is compact.

I'm afraid that we are going to want this theorem with $I = H_1$. So there are no real short cuts, and even if you are willing to take the theorem on trust for the moment, you are going to have to think about continuous functions and compact sets in non-trivial topological spaces.

2C Proof of 1G I repeat the target:

Kirszbraun's Theorem Let H_1 and H_2 be Hilbert spaces, $A \subseteq H_1$ a set and $f : A \to H_2$ a function. Suppose that $\gamma \ge 0$ is such that $\|f(x) - f(y)\| \le \gamma\|x - y\|$ for all x, $y \in A$. Then there is a function $g : H_1 \to H_2$ such that $g(x) = f(x)$ for every $x \in A$ and $\|g(x) - g(y)\| \le \gamma\|x - y\|$ for all x, $y \in H_1$.

proof (a) If $A = \emptyset$ then we can just set $g(x) = 0$ for every $x \in H_1$. So let us suppose from now on that A has at least one member; fix $a \in A$.

(b) For each $x \in H_1$, let $B_x \subseteq H_2$ be the ball

$$\{y : \|y\| \le \|f(a)\| + \gamma\|x - a\|\}$$

so that B_x is a compact subset of H_2 when given its weak topology. By Tychonoff's theorem, $X = \prod_{x \in H_1} B_x$ is compact in its product topology. Now, for any finite set $I \subseteq H_1$, set

$$\begin{aligned} F_I = \{g : g \in X,\ &g(x) = f(x) \text{ for every } x \in I \cap A, \\ &\|g(x) - g(y)\| \le \gamma\|x - y\| \text{ for every } x,\ y \in I\}. \end{aligned}$$

(c) We have to check that all these sets F_I are non-empty. To see this, note that by Corollary 1F there is a function $g_0 : I \to H_2$ such that $g_0(x) = f(x)$ for every $x \in (I \cup \{a\}) \cap A$ and $\|g_0(x) - g_0(y)\| \le \gamma\|x - y\|$ for every x, $y \in I \cup \{a\}$. Now

$$\begin{aligned} \|g_0(x)\| &= \|f(a) + (g(x) - f(a))\| \le \|f(a)\| + \|g(x) - f(a)\| \\ &= \|f(a)\| + \|g_0(x) - g_0(a)\| \le \|f(a)\| + \gamma\|x - a\|, \end{aligned}$$

so $g_0(x) \in B_x$, for every $x \in I$. If we now set

$$\begin{aligned} g(x) &= g_0(x) \text{ for } x \in I, \\ &= 0 \text{ for other } x \in H_1, \end{aligned}$$

we shall have $g \in X$ and therefore $g \in F_I$.

(d) We have to check that all the sets F_I are closed. To see this, note that for each $x \in I \cap A$ the map $g \mapsto g(x) : X \to B_x$ is continuous and the set $\{f(x)\}$ is a closed set, so $\{g : g \in X,\ g(x) = f(x)\}$ is closed. At the same time, for $x,\ y \in I$, the maps $g \mapsto g(x)$ and $g \mapsto g(y)$ from X to H_2 are continuous when H_2 is given its weak topology; while the function $z \mapsto (z|c) : H_2 \to \mathbb{R}$ is continuous for every $c \in H_2$. So $g \mapsto (g(x)|c)$, $g \mapsto (g(y)|c)$ and $g \mapsto (g(x) - g(y)|c)$ from X to $\mathbb{R}$ are all continuous; consequently $\{g : (g(x) - g(y)|c) \le \gamma\|x - y\|\}$ is closed in X. Now by Lemma 1Ca,

$$\{g : \|g(x) - g(y)\| \le \gamma\|x - y\|\} = \bigcap_{\|c\|\le 1}\{g : (g(x) - g(y)|c) \le \gamma\|x - y\|\}$$

is an intersection of closed subsets of X, so is closed. Finally,

$$F_I = \bigcap_{x\in I\cap A}\{g : g(x) = f(x)\} \cap \bigcap_{x,y\in I}\{g : \|g(x) - g(y)\| \le \gamma\|x - y\|\}$$

is in turn an intersection of closed sets, so is closed.

(e) Set $\mathcal{E} = \{F_I : I \subseteq H_1 \text{ is finite}\}$. Then we have just seen that $\mathcal{E}$ is a family of closed sets. Also it has the finite intersection property. For if $I_0, \dots , I_n \subseteq H_1$ are finite sets containing a, then $I = \bigcup_{i\le n} I_i$ is finite, and

$$\bigcap_{i\le n} F_{I_i} \supseteq F_I \ne \emptyset.$$

(f) Thus we have a compact space X and a family $\mathcal{E}$ of closed subsets of X with the finite intersection property. There is therefore a $g \in X$ which belongs to every member of $\mathcal{E}$, that is, $g \in F_I$ whenever $I \subseteq H_1$ is finite and contains a. In particular, if $x \in A$, then $g \in F_{\{a,x\}}$ so $g(x) = f(x)$, while if $x,\ y \in H_1$, then $g \in F_{\{a,x,y\}}$ so $\|g(x) - g(y)\| \le \gamma\|x - y\|$. But this means that g is a function with the properties we need.

3 Second proof

The argument in 2C above took a variety of topological concepts and manipulations for granted. Even if you are interested only in the case $H_1 = H_2 = \mathbb{R}^2$ (so that all the sets B_x are just disks), and have some idea of what a compact set in the plane looks like, and are willing to take Tychonoff's theorem on trust, the proof may have gone rather briskly. I will therefore describe an alternative route. It will take longer because I will not call explicitly on either the Banach-Alaoglu theorem or Tychonoff's theorem, and in fact I will incorporate what amount to proofs of these theorems. In a sense, therefore, this is a more 'elementary' proof, and may be more accessible if you have done no functional analysis. Even if you have done enough to make §2 reasonably straightforward, the techniques here are an essential part of general topology, and are worth studying for that reason.

3A Filters: three definitions (a) If X is a set, a **filter** on X is a family $\mathcal{F}$ of subsets of X such that
$X \in \mathcal{F}$;
$\emptyset \notin \mathcal{F}$;
if $A \in \mathcal{F}$ and $A \subseteq B \subseteq X$ then $B \in \mathcal{F}$;
if $A,\ B \in \mathcal{F}$ then $A \cap B \in \mathcal{F}$.

(b) If X is a set, an **ultrafilter** on X is a filter $\mathcal{G}$ on X such that
for every $A \subseteq X$, either $A \in \mathcal{G}$ or $X \setminus A \in \mathcal{G}$.

(c) If X is a set, a **filter base** on X is a family $\mathcal{E}$ of subsets of X such that
$\mathcal{E}$ is not empty;
$\emptyset \notin \mathcal{E}$;
if $A,\ B \in \mathcal{E}$ there is a $C \in \mathcal{E}$ such that $C \subseteq A \cap B$.

(d) If you have never seen these words before, don't worry; I will try to explain everything fully. But you will need to watch very closely, because we shall be completely dependent on the logic of set theory, and every word of the definitions above is vital. Perhaps I should point out at once that if $\mathcal{F}$ is a filter on X then X cannot be empty, because $X \in \mathcal{F}$ and $\emptyset \notin \mathcal{F}$.

(e) We can get a bit more practice with the following: if X is a set and $\mathcal{E}$ is a filter base on X, there is a filter $\mathcal{F}$ on X which includes $\mathcal{E}$. In fact there is an easy formula for one. Try

$$\mathcal{F} = \{A : A \subseteq X \text{ and there is some } E \in \mathcal{E} \text{ such that } E \subseteq A\}.$$

Working through the definition of 'filter', we have

— every member of $\mathcal{F}$ is a subset of X because that's written into the formula for $\mathcal{F}$;

— $\emptyset \notin \mathcal{F}$ because the only subset of $\emptyset$ is itself, and $\emptyset \notin \mathcal{E}$;

— $X \in \mathcal{F}$ because $\mathcal{E}$ is not empty, and if $A \in \mathcal{E}$ then $A \subseteq X \subseteq X$;

— if $A \in \mathcal{F}$ and $A \subseteq B \subseteq X$, there is an $E \in \mathcal{E}$ such that $E \subseteq A$, and now $E \subseteq B$, so $B \in \mathcal{F}$;

— if A, $B \in \mathcal{F}$ then there are E, $F \in \mathcal{E}$ such that $E \subseteq A$ and $F \subseteq B$; now there is a $C \in \mathcal{E}$ such that $C \subseteq E \cap F$, so we have $C \subseteq A \cap B \subseteq X$ and $A \cap B \in \mathcal{F}$.

So all the clauses of the definition in (a) are satisfied, mostly because they match the clauses of the definition in (c).

3B Limits along filters The most important reason for thinking about filters is the following. Suppose that X is a set, $\mathcal{F}$ is a filter on X, $\phi : X \to \mathbb{R}$ is a function, and $\alpha \in \mathbb{R}$. Then we say that ϕ **converges to** α **along** $\mathcal{F}$ if for every $\epsilon > 0$ the set $\{t : t \in X,\, |\phi(t) - \alpha| \le \epsilon\}$ belongs to $\mathcal{F}$.

I think this is the first ϵ in this note. We got through the whole of the first proof of Kirszbraun's theorem without a single ϵ, though I have a limit in part (c-iii) of the proof of Lemma 1C, and all the ideas on continuity in 2A-2C depend, at bottom, on ϵs and δs.

We now have some more practice on elementary properties of filters. These ought to remind you of limits of sequences.

3C Proposition Let X be a set and $\mathcal{F}$ a filter on X.

(a) If $\phi : X \to \mathbb{R}$ is a function, there can be at most one $\alpha \in \mathbb{R}$ such that ϕ converges to α along $\mathcal{F}$. We can therefore write

$$\lim_{t\to\mathcal{F}} \phi(t) = \alpha$$

to mean that this happens.

(b) Suppose that $\phi : X \to \mathbb{R}$ is a function and $\lim_{t\to\mathcal{F}} \phi(t) = \alpha$. Then $\lim_{t\to\mathcal{F}} \beta\phi(t) = \beta\alpha$ for every $\beta \in \mathbb{R}$.

(c) If $\phi : X \to \mathbb{R}$ and $\psi : X \to \mathbb{R}$ are such that $\lim_{t\to\mathcal{F}} \phi(t) = \alpha$ and $\lim_{t\to\mathcal{F}} \psi(t) = \beta$, then $\lim_{t\to\mathcal{F}} \phi(t) + \psi(t) = \alpha + \beta$.

proof (a) If ϕ converges along $\mathcal{F}$ to both α and α', then for any $\epsilon > 0$ the sets

$$A = \{t : t \in X,\, |\phi(t) - \alpha| \le \frac{1}{2}\epsilon\}, \quad A' = \{t : t \in X,\, |\phi(t) - \alpha'| \le \frac{1}{2}\epsilon\}$$

both belong to $\mathcal{F}$. So $A \cap A' \in \mathcal{F}$ and $A \cap A'$ is not empty, by the fourth and second clauses in the definition of 'filter'. Take any $t \in A \cap A'$; then

$$|\alpha - \alpha'| \le \|\alpha - \phi(t)| + |\phi(t) - \alpha'| \le \frac{1}{2}\epsilon + \frac{1}{2}\epsilon = \epsilon.$$

So $|\alpha - \alpha'| \le \epsilon$ for every $\epsilon > 0$ and $\alpha = \alpha'$.

(b) Let $\epsilon > 0$. Then the set

$$A = \{t : t \in X,\, |\phi(t) - \alpha| \le \tfrac{\epsilon}{1+|\beta|}\}$$

belongs to $\mathcal{F}$. Now if $t \in A$,

$$|\beta\phi(t) - \beta\alpha| = |\beta||\phi(t) - \alpha| \le \tfrac{\epsilon|\beta|}{1+|\beta|} \le \epsilon.$$

So $\{t : t \in X,\, |\beta\phi(t) - \beta\alpha| \le \epsilon\}$ includes A and belongs to $\mathcal{F}$. As ϵ is arbitrary, $\lim_{t\to\mathcal{F}} \beta\phi(t) = \beta\alpha$.

(c) Let $\epsilon > 0$. Then the sets

$$A = \{t : t \in X,\, |\phi(t) - \alpha| \le \frac{\epsilon}{2}, \quad A = \{t : t \in X,\, |\psi(t) - \beta| \le \frac{\epsilon}{2}\}$$

belong to $\mathcal{F}$. Now if $t \in A \cap B$,

$$|\phi(t) + \psi(t)) - \alpha - \beta| \le |\phi(t) - \alpha| + |\psi(t) - \beta| \le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

So $\{t : t \in X, |(\phi(t)+\psi(t))-(\alpha+\beta)| \le \epsilon\}$ includes $A \cap B$ and belongs to $\mathcal{F}$. As ϵ is arbitrary, $\lim_{t\to\mathcal{F}} \phi(t)+\psi(t) = \alpha+\beta$.

Remark You will see that the definition of 'filter' is almost exactly what is needed to make this proposition work. (I haven't used the clause '$X \in \mathcal{F}$', but everything else is called on.)

Not everyone who teaches the elementary theory of convergence of sequences is careful to point out that a sequence can have at most one limit. But I hope you can see that the grammar of the sentences '$\lim_{n\to\infty} t_n = \alpha$', '$\lim_{t\to\mathcal{F}} \phi(t) = \alpha$' insists on this being true. Of course not all sequences have limits, and to have a limit along a filter is also the exception rather than the rule – any appearance to the contrary is because we don't trouble to mention the cases when there is no limit.

3D Now for something which doesn't correspond directly to anything you will have seen for limits of sequences.

Proposition Let X be a set, $\mathcal{G}$ an ultrafilter on X and $\phi : X \to \mathbb{R}$ a function such that $B = \{t : t \in X, |\phi(t)| \le \gamma\}$ belongs to $\mathcal{G}$ for some $\gamma \ge 0$. Then $\lim_{t\to\mathcal{G}} \phi(t)$ is defined and belongs to $[-\gamma, \gamma]$.

proof For $\alpha \in \mathbb{R}$, set

$$A_\alpha = \{t : t \in X, \phi(t) \ge \alpha\}.$$

Consider $C = \{\alpha : \alpha \in \mathbb{R}, A_\alpha \in \mathcal{G}\}$.

We have

$$A_{-\gamma} \supseteq B \in \mathcal{G},$$

so $-\gamma \in C$.

If $\alpha \in C$ then $A_\alpha \cap B \in \mathcal{G}$ and $A_\alpha \cap B \ne \emptyset$. There is therefore a $t \in X$ such that $\phi(t) \ge \alpha$ and $|\phi(t)| \le \gamma$; in which case $\alpha \le \phi(t) \le \gamma$. Thus γ is an upper bound for C.

Putting these together, C is a non-empty subset of $\mathbb{R}$ with an upper bound in $\mathbb{R}$, and has a supremum β say, with $-\gamma \le \beta \le \gamma$.

If $\epsilon > 0$, then $\beta + \epsilon \notin C$, so that $A_{\beta+\epsilon} \notin \mathcal{G}$. But $\mathcal{G}$ is supposed to be an ultrafilter, so $X \setminus A_{\beta+\epsilon} \in \mathcal{G}$. At the same time, $\beta - \epsilon$ is not an upper bound for C, so there is an $\alpha \in C$ such that $\beta - \epsilon \le \alpha$, and $A_\alpha \in \mathcal{G}$. Putting these together,

$$A_\alpha \setminus A_{\beta+\epsilon} = A_\alpha \cap (X \setminus A_{\beta+\epsilon}) \in \mathcal{G}.$$

But if $t \in A_\alpha \setminus A_{\beta+\epsilon}$, then

$$\beta - \epsilon \le \alpha \le \phi(t) \le \beta + \epsilon$$

and $|\phi(t) - \beta| \le \epsilon$. Thus $\{t : t \in X, |\phi(t) - \beta| \le \epsilon\}$ includes $A_\alpha \setminus A_{\beta+\epsilon}$ and must belong to $\mathcal{G}$. As ϵ is arbitrary, $\lim_{t\to\mathcal{G}} \phi(t) = \beta$ is defined and belongs to $[-\gamma, \gamma]$.

3E The Ultrafilter Lemma Whenever $\mathcal{F}$ is a filter on a set X, there is an ultrafilter $\mathcal{G}$ on X including $\mathcal{F}$.

Remark I will make no attempt to prove this, for reasons which I will try to explain in §4. For the moment, I will ask you to take it as a theorem of more or less the same kind as Tychonoff's theorem (2B above). See `http://en.wikipedia.org/wiki/Ultrafilter_lemma`.

3F Proof of 1G Once again, I repeat the target:

Kirszbraun's Theorem Let H_1 and H_2 be Hilbert spaces, $A \subseteq H_1$ a set and $f : A \to H_2$ a function. Suppose that $\gamma \ge 0$ is such that $\|f(x) - f(y)\| \le \gamma\|x - y\|$ for all $x, y \in A$. Then there is a function $g^* : H_1 \to H_2$ such that $g^*(x) = f(x)$ for every $x \in A$ and $\|g^*(x) - g^*(y)\| \le \gamma\|x - y\|$ for all $x, y \in H_1$.

proof I begin by setting up just the same structure as that used in 2C.

(a) If $A = \emptyset$ then we can set $g(x) = 0$ for every $x \in H_1$. So let us suppose from now on that A has at least one member; fix $a \in A$.

For each $x \in H_1$, let $B_x \subseteq H_2$ be the ball

$$\{y : \|y\| \leq \|f(a)\| + \gamma\|x - a\|\}.$$

Now, for any finite set $I \subseteq H_1$, set

$$F_I = \{g : g \in X,\ g(x) = f(x) \text{ for every } x \in I \cap A,$$
$$\|g(x) - g(y)\| \leq \gamma\|x - y\| \text{ for every } x,\ y \in I\}.$$

(b) We have to check that all these sets F_I are non-empty. To see this, note that by Corollary 1F there is a function $g_0 : I \to H_2$ such that $g_0(x) = f(x)$ for every $x \in (I \cup \{a\}) \cap A$ and $\|g_0(x) - g_0(y)\| \leq \gamma\|x - y\|$ for every $x,\ y \in I \cup \{a\}$. Now

$$\|g_0(x)\| = \|f(a) + (g(x) - f(a))\| \leq \|f(a)\| + \|g(x) - f(a)\|$$
$$= \|f(a)\| + \|g_0(x) - g_0(a)\| \leq \|f(a)\| + \gamma\|x - a\|,$$

so $g_0(x) \in B_x$, for every $x \in I$. If we now set

$$g(x) = g_0(x) \text{ for } x \in I,$$
$$= 0 \text{ for other } x \in H_1,$$

we shall have $g \in X$ and therefore $g \in F_I$.

(c) Now we start to diverge from the line in 2C. If $I,\ J \subseteq H_1$ are finite sets then $I \cup J$ is a finite subset of H_1 and $F_{I \cup J} \subseteq F_I \cap F_J$. So $\mathcal{E} = \{F_I : I \subseteq H_1 \text{ is finite}\}$ is a filter base on X. There are therefore a filter $\mathcal{F}$ on X including $\mathcal{E}$, by 3Ae above, and an ultrafilter $\mathcal{G}$ on X including $\mathcal{F}$, by the Ultrafilter Lemma.

(d) Suppose that $x \in H_1$ and $z \in H_2$. Then

$$|(f(x)|z)| \leq \|f(x)\|\|z\|$$

(Lemma 3(a-i))

$$\leq (\|f(a)\| + \gamma\|x - a\|)\|z\|$$

for every $f \in X$. By Proposition 3D, $\lim_{f\to\mathcal{G}}(f(x)|z)$ is defined; call it $\phi_x(z)$. We also have $|\phi_x(z)| \leq (\|f(a)\| + \gamma\|x - a\|)\|z\|$.

(e) Fix $x \in H_1$ for the moment. Then we have a functional $\phi_x : H_2 \to \mathbb{R}$. This is linear, because by Lemma 3Cc and 3Cd,

$$\phi_x(\alpha z) = \lim_{f\to\mathcal{G}}(f(x)|\alpha z) = \lim_{f\to\mathcal{G}}\alpha(f(x)|z)$$
$$= \alpha\lim_{f\to\mathcal{G}}(f(x)|z) = \alpha\phi_x(z),$$

$$\phi_x(w + z) = \lim_{f\to\mathcal{G}}(f(x)|w + z) = \lim_{f\to\mathcal{G}}(f(x)|w) + (f(x)|z)$$
$$= \lim_{f\to\mathcal{G}}(f(x)|w) + \lim_{f\to\mathcal{G}}(f(x)|z) = \phi_x(w) + \phi_x(z)$$

for all $w,\ z \in H_2$ and $\alpha \in \mathbb{R}$. We have already seen, at the end of (d) just above, that for any $x \in H_1$ there is a constant $\gamma_x = \|f(a)\| + \gamma\|x - a\|$ such that $|\phi_x(z)| \leq \gamma_x$ whenever $z \in H_2$ and $\|z\| \leq 1$. So Lemma 1Cd tells us that there is a unique member of H_2, which we can call $g^*(x)$, such that $\phi_x(z) = (g^*(x)|z)$ for every $z \in H_2$.

(f) This defines a function g^* from H_1 to H_2. Now consider its properties. Suppose that $x \in A$. Then $F_{\{x\}} \in \mathcal{E} \subseteq \mathcal{F} \subseteq \mathcal{G}$, and $g(x) = f(x)$ for every $g \in F_{\{x\}}$. If $z \in H_2$ and $\epsilon > 0$, then $\{g : |(g(x)|z) - \phi_x(z)| \leq \epsilon\} \in \mathcal{G}$, so $F_{\{x\}} \cap \{g : |(g(x)|z) - \phi_x(z)| \leq \epsilon\}$ belongs to $\mathcal{G}$ and is not empty; take $g \in F_{\{x\}}$ such that $|(g(x)|z) - \phi_x(z)| \leq \epsilon$. Then

$$|(f(x) - g^*(x)|z)| = |(f(x)|z) - (g^*(x)|z)| = |(g(x)|z) - \phi_x(z)| \leq \epsilon.$$

As ϵ is arbitrary, $(f(x) - g^*(x)|z) = 0$. This is true for every $z \in H_2$, so $f(x) = g^*(x)$.

(g) Now suppose that $x, y \in H_1$. Then $F_{\{x,y\}} \in \mathcal{E} \subseteq \mathcal{G}$, and $\|g(x)-g(y)\| \le \gamma\|x-y\|$ for every $g \in F_{\{x,y\}}$. Again, take any $z \in H_2$ and $\epsilon > 0$. We know that

$$\lim_{g\to\mathcal{G}}(g(x)|z) = \phi_x(z), \quad \lim_{g\to\mathcal{G}}(g(y)|z) = \phi_y(z),$$

so

$$\lim_{g\to\mathcal{G}}(g(x)-g(y)|z) = \lim_{g\to\mathcal{G}}(g(x)|z)-(g(y)|z) = \lim_{g\to\mathcal{G}}(g(x)|z) - \lim_{g\to\mathcal{G}}(g(y)|z)$$

(using 3Cb and 3Cc again)

$$= \phi_x(z) - \phi_y(z) = (g^*(x)|z) - (g^*(y)|z) = (g^*(x)-g^*(y)|z).$$

Accordingly

$$\{g : |(g(x)-g(y)|z) - (g^*(x)-g^*(y)|z)| \le \epsilon\}$$

belongs to $\mathcal{G}$ and must meet $F_{\{x,y\}}$. Take $g \in F_{\{x,y\}}$ such that $|(g(x)-g(y)|z)-(g^*(x)-g^*(y)|z)| \le \epsilon$; then

$$|(g^*(x)-g^*(y)|z)| \le \epsilon + |(g(x)-g(y)|z)| \le \epsilon + \|g(x)-g(y)\|\|z\|$$

(1C(a-i) again)

$$\le \epsilon + \gamma\|x-y\|\|z\|.$$

As ϵ is arbitrary, $|(g^*(x)-g^*(y)|z)| \le \gamma\|x-y\|\|z\|$; as z is arbitrary, $\|g^*(x)-g^*(y)\| \le \gamma\|x-y\|$, by 1C(a-ii).

(h) Putting (f) and (g) together, we see that g^* has the required properties, and the proof is complete.

4 The magic of choice

In the proofs above, I have mentioned three major results: a special case of the Banach-Alaoglu theorem (2A), Tychonoff's theorem (2B) and the Ultrafilter Lemma (3E). All three can be proved if we are willing to use the Axiom of Choice (`http://en.wikipedia.org/wiki/Axiom_of_choice`). Here I should like to offer some thoughts on what we should make of this.

4A Miracles and magic (a) I called 1B a 'miracle'. This is pitching it strong, but conveys the notion that there is something singular about it. In distinction to this, I would call the Ultrafilter Lemma 'magic'. We intone the spell 'let $\mathcal{G}$ be an ultrafilter including $\mathcal{F}$' and a door opens. But $\mathcal{G}$ is indescribable and uncontrolled. I didn't offer any actual examples of ultrafilters because *without* a special axiom of some kind there are very few of them, and all those which can be described explicitly in the ordinary language of mathematics are quite useless. An ultrafilter is generally a jinn which we have called into our service. If we call at the right moment, we can get very good service. But is this black magic or white magic?

(b) The Axiom of Choice can be stated in various ways, but the one which seems most natural in the context here is

> (AC) If $\langle X_i\rangle_{i\in I}$ is any family of non-empty sets, then $\prod_{i\in I} X_i$ is not empty, that is, there is a function g defined on the set I such that $g(i) \in X_i$ for every $i \in I$.

Put like this, AC seems an entirely natural principle. For finite sets I, in fact, it is easily proved by induction on $\#(I)$. I actually used a simple version of it in the proof of Lemma 1C above. In part (c-ii) of that proof, I wrote 'since every C_n is non-empty, we can choose a sequence $\langle z_n\rangle_{n\in\mathbb{N}}$ such that $z_n \in C_n$ for every n', that is, $\langle z_n\rangle_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} C_n$. I do not think I have ever had a student who noticed that anything special was being done in such cases (I am sure I did not notice myself, until it was pointed out to me).

The proofs that AC implies the Ultrafilter Lemma and Tychonoff's theorem are not natural in the way that choosing a sequence is, but are undoubtedly part of ordinary abstract pure mathematics. It is fair to say that the great majority of pure mathematicians use AC and its consequences without scruple whenever they come in handy.

(c) I myself believe that it is better to be conscious of the occasions on which one is relying on it. Partly I think that one of the things mathematics is for is to raise our consciousness of what we are really doing.

Partly I feel that proofs not using choice (when available) are more illuminating than the alternatives, even when they are longer and more effort. And partly I believe that there are rival axioms, actually contradicting AC, which are well worth exploring (the most interesting, at the moment, being the 'Axiom of Determinacy', http://en.wikipedia.org/wiki/Axiom_of_determinacy). You will be quite unable to join in this exploration if you are not aware which of your favourite theorems may have vanished in the new landscape.

(d) So we are led naturally to the question: do we really need AC to prove Kirszbraun's theorem? Especially we should ask: do we really need AC to prove Kirszbraun's theorem when both Hilbert spaces are finite-dimensional Euclidean spaces, starting with $\mathbb{R}^2$? because such strong magic, whether black or white, should not be needed if we can see what we are doing.

4B First things first (a) As I noted in 4Ab, I did use AC, in a very simple form, during the proof of 1Cc. In fact it is not really necessary here. The key to avoiding it is to use the correct definition of 'Hilbert space'. Now everyone agrees that a Hilbert space is a complete inner product space (http://en.wikipedia.org/wiki/Hilbert_space). But not everyone uses the correct definition of 'complete'. The standard definition (http://en.wikipedia.org/wiki/Complete_metric_space) declares that a metric space (X, ρ) is complete if every Cauchy sequence in X is convergent. Working from this, you won't be able to prove further properties of X without getting hold of a Cauchy sequence, which is what I do in part (c-ii) of the proof of 1C. And to get an actual sequence you are likely to have to choose all its terms simultaneously – or, what demands a slightly stronger use of AC, one at a time in an inductive process. So I myself use a different definition of 'complete metric space'. I say that a metric space (X, ρ) is complete if every Cauchy *filter* converges. I see that I have to give you two more definitions. First, if (X, ρ) is a metric space, a filter $\mathcal{F}$ on X **converges** to a point $x \in X$ if $\{y : y \in X,\ \rho(x, y) \le \epsilon\}$ belongs to $\mathcal{F}$ for every $\epsilon > 0$. (This is like 3B, but without the function ϕ – or, if you like, replacing ϕ with the identity function from X to itself.) Second, a filter $\mathcal{F}$ on X is a **Cauchy** filter if for every $\epsilon > 0$ there is a set $A \in \mathcal{F}$ with diameter at most ϵ, that is, such that $\rho(x, y) \le \epsilon$ for all x, $y \in A$.

(b) If you will follow me this far, and agree that a Hilbert space is an inner product space in which every Cauchy filter converges, then the proof of 1Cc can be rewritten with no choosing of sequences. For the sets $C_n = \{z : z \in C,\ \|z - b\| \le \alpha + 4^{-n}\}$ are defined by a formula. I'm not asking a jinn to tell me what to do here. Next, $C_m \cap C_n = C_n$ whenever $m \le n$, so $\mathcal{E} = \{C_n : n \in \mathbb{N}\}$ is a filter base. There is therefore a filter $\mathcal{F}$ on H including $\mathcal{E}$, by 3Ae (which also gives an explicit formula for $\mathcal{F}$). Next, the diameter of C_n is at most $2^{-n}\sqrt{8\alpha + 4}$ (part (c-i) of the proof of 1C). Since $\lim_{n\to\infty} 2^{-n}\sqrt{8\alpha + 4} = 0$, and every C_n belongs to $\mathcal{F}$, $\mathcal{F}$ is a Cauchy filter. $\mathcal{F}$ therefore converges to some $b' \in H$. Of course we need to check that because $C \in \mathcal{F}$ and C is closed, $b' \in C$, and also that $\|b' - b\| \le \alpha$, which takes a little manoeuvre of the same kind as those in parts (f) and (g) of the proof in 3F. But now we can repeat the argument of part (c-iii) of the proof of 1C.

(c) Obviously this won't do unless the Hilbert spaces we know and love, and also those which turn up in other parts of mathematics (either pure or applied), are complete in the Cauchy-filter sense. But they all are. For instance, there is a proof that a Cauchy filter on $\mathbb{R}$ is convergent which essentially follows the line of Proposition 3D above. We just have to find a different reason for concluding that if $A_{\beta+\epsilon} \notin \mathcal{F}$ then $X \setminus A_{\beta+2\epsilon} \in \mathcal{F}$. (Look at where a set of diameter at most ϵ which belongs to $\mathcal{F}$ must be.) And now we can get to finite-dimensional Euclidean spaces, and even ℓ^2, by looking at coordinates in the same way as we do for Cauchy-sequence completeness. We need slightly more refined techniques, but the hardest part of the exercise is getting used to handling filters at all. It is trickier to get one's mind around the idea of filter than the idea of sequence. But, in return, we can write down our own formulae for filters in contexts where we have to ask for magical help in building sequences.

4C The heavy lifting All the rest of §1 goes through with no call for AC. By the time we have reached 1F, we are a little short of explicit formulae, but we are claiming the existence of only one function g at a time, and that function defined only on a finite set $I \cup J$. (If we wanted to talk about a whole family $\langle g_{IJ}\rangle_{I,J\subseteq H_1 \text{ are finite}}$, of course, we should then need some form of AC.) Moreover, if (as I have repeatedly said) we care mostly about the finite-dimensional case, then there is no problem in 2A; when $H = \mathbb{R}^n$, the weak topology is actually the norm topology, and we have direct proofs to show that closed balls are compact. (I must say I think these are easier and cleaner if you use filters, but you don't have to. You're allowed to induce on n, if you find it helps. Just make sure that you aren't choosing sequences without giving formulae for them.)

However, even if $H_1 = \mathbb{R}$ and $H_2 = \mathbb{R}^2$, the argument of 2C demands that we look at an uncountably infinite product $X = \prod_{x\in H_1} B_x$. And this really does seem to involve us with Tychonoff's theorem, by no means in its full strength (because all the compact sets B_x are simple ones, by the standards of general topology), but certainly in a form which isn't consistent with the Axiom of Determinacy, for instance.

I have not figured out just how much of the Ultrafilter Lemma we need for Kirszbraun's theorem in the full generality stated. (We don't need the whole of AC; it's known that it's possible for AC to be false but the Ultrafilter Lemma true – see Halpern & Levy 71.) However, for *separable* H_1 and H_2 (including Euclidean spaces and ℓ^2, and the great majority of Hilbert spaces required in ordinary applications), there is a way round, necessarily harder work with no jinn to help us, but a good example of technique.

4D Clearing the way I had better make it clear what I mean by saying that a Hilbert space H is 'separable'; I mean that there is a countable set which is dense for the norm. Since a countable set is, by definition, either finite or equipollent with $\mathbb{N}$, and since H is not empty, there will be a sequence $\langle x_n\rangle_{n\in\mathbb{N}}$ in H such that $\{x_n : n\in\mathbb{N}\}$ is dense.

4E Next, it will be helpful to have the notion of 'limit superior' along filters.

Definition If X is a set, $\mathcal{F}$ is a filter on X and $\phi : X \to \mathbb{R}$ is a function, then I will write

$$\limsup_{t\to\mathcal{F}} \phi(t) = \inf_{A\in\mathcal{F}} \sup_{t\in A} \phi(t).$$

In this formula, I am willing to allow $\sup_{t\in A}\phi(t) = \infty$ if ϕ is not bounded above on A, $\limsup_{t\to\mathcal{F}}\phi(t) = \infty$ if ϕ is unbounded above on every $A\in\mathcal{F}$, and $\limsup_{t\to\mathcal{F}}\phi(t) = -\infty$ if $\{\sup_{t\in A}\phi(t) : A\in\mathcal{F}\}$ is unbounded below.

4F Proposition Let X be a set and $\mathcal{F}$ a filter on X. Suppose that $\phi : X\to\mathbb{R}$ is a function such that for every $\epsilon > 0$ there is a function $\psi : X\to\mathbb{R}$ such that (α) $\lim_{t\to\mathcal{F}}\psi(t)$ is defined in $\mathbb{R}$ (β) there is an $A\in\mathcal{F}$ such that $|\phi(t)-\psi(t)|\le\epsilon$ for every $t\in A$. Then $\lim_{t\to\mathcal{F}}\phi(t)$ is defined in $\mathbb{R}$.

proof (a) There are a $B_0\in\mathcal{F}$ and a $\psi_0 : X\to\mathbb{R}$ such that $\beta_0 = \lim_{t\to\mathcal{F}}\psi_0(t)$ is defined and $|\phi(t)-\psi_0(t)|\le 1$ for every $t\in B_0$. Next, there is a $B_1\in\mathcal{F}$ such that $|\psi_0(t)-\beta_0|\le 1$ for every $t\in B_1$. In this case, $B_0\cap B_1\in\mathcal{F}$ and $|\phi(t)-\beta_0|\le 2$ for every $t\in B_0\cap B_1$. It follows at once that

$$\limsup_{t\to\mathcal{F}}\phi(t)\le\sup_{t\in B_0\cap B_1}\phi(t)\le\beta_0+2;$$

at the same time, if A is any member of $\mathcal{F}$, there is a $t_0\in A\cap B_0\cap B_1$, so

$$\sup_{t\in A}\phi(t)\ge\phi(t_0)\ge\beta_0-2;$$

putting these together, $\alpha = \limsup_{t\to\mathcal{F}}\phi(t)$ lies between β_0-2 and β_0+2, and is finite.

(b) Let $\epsilon>0$. Then there is an $A_0\in\mathcal{F}$ such that $\sup_{t\in A_0}\phi(t)\le\alpha+\epsilon$; there are a $\psi : X\to\mathbb{R}$ and an $A_1\in\mathcal{F}$ such that $\beta = \lim_{t\to\mathcal{F}}\psi(t)$ is defined and $|\phi(t)-\psi(t)|\le\frac{1}{5}\epsilon$ for every $t\in A_1$; and there is an $A_2\in\mathcal{F}$ such that $|\psi(t)-\beta|\le\frac{1}{5}\epsilon$ for every $t\in A_2$. As $A = A_0\cap A_1\cap A_2$ belongs to $\mathcal{F}$, $\sup_{t\in A}\phi(t)\ge\alpha$, and there is a $t_1\in A$ such that $\phi(t_1)\ge\alpha-\frac{1}{5}\epsilon$; in which case

$$\beta\ge\psi(t_1)-\frac{1}{5}\epsilon\ge\phi(t_1)-\frac{2}{5}\epsilon\ge\alpha-\frac{3}{5}\epsilon.$$

Now, for any $t\in A$,

$$\alpha+\epsilon\ge\phi(t)\ge\psi(t)-\frac{1}{5}\epsilon\ge\beta-\frac{2}{5}\epsilon\ge\alpha-\epsilon,$$

that is, $|\phi(t)-\alpha|\le\epsilon$. As ϵ is arbitrary, $\lim_{t\to\mathcal{F}}\phi(t) = \alpha$ is defined.

4G Lemma (a) Let (X,ρ) be a complete separable metric space, and $F\subseteq X$ a closed subset. Then F is separable.

(b) Let (X, ρ) be a metric space, (Y, σ) a complete metric space, A a subset of X, and $f : A \to Y$ a function such that $\sigma(f(x), f(x')) \le \gamma\rho(x, x')$ for all x, $x' \in A$. Then there is a unique function $g : \overline{A} \to Y$ such that $g(x) = f(x)$ for every $x \in A$ and $\sigma(g(x), g(x')) \le \gamma\rho(x, x')$ for all x, $x' \in \overline{A}$.

proof (a) If X is empty, this is trivial. Otherwise, let D be a countable dense subset of X and $\langle x_n\rangle_{n\in\mathbb{N}}$ a sequence running over D. Set $K = \{(n, k) : n, k \in \mathbb{N}, U(x_n, 2^{-k})\cap F \ne \emptyset\}$, where $U(x, \epsilon) = \{x' : \rho(x', x) < \epsilon\}$ for $x \in X$ and $\epsilon > 0$. For $(n, k) \in K$, define $\langle z_{nki}\rangle_{i\in\mathbb{N}}$ inductively by saying that

$z_{nk0} = x_n$,

given that $U(z_{nki}, 2^{-k-i}) \cap F \ne \emptyset$, $z_{n,k,i+1} = x_m$ where m is the least member of $\mathbb{N}$ such that $U(z_{nki}, 2^{-k-i}) \cap U(x_m, 2^{-k-i-1}) \cap F \ne \emptyset$.

(To see that there always is such an m, note that there is an $x' \in F \cap U(z_{nki}, 2^{-k-i})$, and now there must be an m such that $\rho(x', x_m) < 2^{-k-i-1}$.) In this case, $\rho(z_{n,k,i+1}, z_{nki}) \le 2^{-k-i} + 2^{-k-i-1}$ for every i, so $\langle z_{nki}\rangle_{i\in\mathbb{N}}$ is a Cauchy sequence and has a limit x'_{nk} in X, because X is complete. Now $\rho(x'_{nk}, z_{nki}) \le 2^{-k-i+1}$ for every i, so $U(x'_{nk}, 2^{-k-i+2}) \supseteq U(z_{nki}, 2^{-k-i})$ meets F, for every i; because F is closed, $x'_{nk} \in F$.

Now $D' = \{x'_{nk} : (n, k) \in K\}$ is a countable subset of F. To see that it is dense in F, take any $x' \in F$ and $k \in \mathbb{N}$. Then there is an $n \in \mathbb{N}$ such that $\rho(x_n, x') < 2^{-k}$ and $(n, k) \in K$. Next, $\rho(x'_{nk}, x_n) = \rho(x'_{nk}, z_{nk0}) \le 2^{-k+1}$, so $\rho(x', x'_{nk}) \le 2^{-k+2}$. As x' and k are arbitrary, D' is dense in F and F is separable.

(b) If A is empty, so is $\overline{A}$, and $g = f$ is the empty function. Otherwise, for each $x \in \overline{A}$,

$$\mathcal{F}_x = \{B : B \subseteq Y, \text{ there is a } \delta > 0 \text{ such that } B \supseteq f[A \cap U(x, \delta)]\}$$

is a Cauchy filter on Y, because $\operatorname{diam}(f[A \cap U(x, \delta)]) \le 2\gamma\delta$ for every $\delta > 0$. If $x \in A$ then $f(x)$ belongs to every member of $\mathcal{F}_x$ so $g(x) = f(x)$. Next, if x, $x' \in \overline{A}$ and $\epsilon > 0$, there are x_1, $x'_1 \in A$ such that $\rho(x, x_1)$, $\rho(x', x'_1)$, $\sigma(g(x), f(x_1))$ and $\sigma(g(x'), f(x'_1))$ are all at most ϵ; now

$$\sigma(g(x), g(x')) \le 2\epsilon + \sigma(f(x_1), f(x'_1)) \le 2\epsilon + \gamma\rho(x_1, x'_1) \le 2\epsilon + \gamma(2\epsilon + \rho(x, x')).$$

This shows that $\sigma(g(x), g(x')) \le \gamma\rho(x, x')$ for all x, $x' \in \overline{A}$. Of course g is unique because it is continuous and agrees with f on a dense subset of $\overline{A}$.

Remark In the hypothesis of part (a), when I wrote that 'X is complete', of course I meant that every Cauchy filter on X is convergent; but it follows easily that every Cauchy sequence converges.

The filters $\mathcal{F}_x$ in part (b) of the proof belong to the standard proof that a uniformly continuous function into a complete Hausdorff uniform space has a uniformly continuous extension; but if you have seen this result only for metric spaces, then it may have been done with Cauchy sequences instead, which are problematic here.

4H Kirszbraun's theorem without AC We now have the following.

Kirszbraun's Theorem Let H_1 be a separable Hilbert space, H_2 a Hilbert space, $A \subseteq H_1$ a set and $f : A \to H_2$ a function. Suppose that $\gamma \ge 0$ is such that $\|f(x) - f(y)\| \le \gamma\|x - y\|$ for all x, $y \in A$. Then there is a function $g^* : H_1 \to H_2$ such that $g^*(x) = f(x)$ for every $x \in A$ and $\|g^*(x) - g^*(y)\| \le \gamma\|x - y\|$ for all x, $y \in H_1$.

proof (a) For most of the proof (down to the end of (e) below) I will suppose that H_2 also is separable. In this case, we can follow the argument of 3F down to the last sentence of part (c) there. We suppose that $a \in A$, that $B_x = \{y : y \in H_2, \|y\| \le \|f(a)\| + \gamma\|x - a\|\}$ for every $x \in H_2$, that $X = \prod_{x\in H_1} B_x$, and that $\mathcal{F}$ is a filter on X containing

$$\begin{aligned} F_I = \{g : g \in X,\ & g(x) = f(x) \text{ for every } x \in I \cap A, \\ & \|g(x) - g(y)\| \le \gamma\|x - y\| \text{ for every } x, y \in I\} \end{aligned}$$

for every finite set $I \subseteq H_1$. Note that the constant function with value 0 belongs to X, so X is not empty. (Of course this is included in the demonstration, in part (b) of the proof in 3F, that $F_\emptyset \ne \emptyset$.)

(b) We no longer expect there to be an ultrafilter including $\mathcal{F}$. But we can build something good enough for our purposes, as follows. Let $D_1 \subseteq H_1$, $D_2 \subseteq H_2$ be countable dense sets. Then $D_1 \times D_2$ is countable, so there is a sequence $\langle(x_n, z_n)\rangle_{n\in\mathbb{N}}$ running over $D_1 \times D_2$. Now define a sequence $\langle\mathcal{G}_n\rangle_{n\in\mathbb{N}}$ of filters on X, as follows. Start with $\mathcal{G}_0 = \mathcal{F}$. Given $\mathcal{G}_n$, note that

$$|(g(x_n)|z_n)| \le \|g(x_n)\|\|z_n\| \le \|f(a)\| + \gamma\|x_n - a\|\|z_n\|$$

for every $g \in X$, so

$$\alpha_n = \limsup_{g\to\mathcal{G}_n}(g(x_n)|z_n)$$

is finite. Set

$$\mathcal{G}_{n+1} = \bigcup_{C\in\mathcal{G}_n,\epsilon>0}\{B : B \subseteq X,\, C \cap \{g : |\alpha_n - (g(x_n)|z_n)| \le \epsilon\} \subseteq B\}.$$

It is easy to see that $B \cap B' \in \mathcal{G}_{n+1}$ whenever B, $B' \in \mathcal{G}_{n+1}$ and that $\emptyset \notin \mathcal{G}_{n+1}$, so that $\mathcal{G}_{n+1}$ is a filter on X including $\mathcal{G}_n$. Continue. At the end of the induction, set $\mathcal{G} = \bigcup_{n\in\mathbb{N}}\mathcal{G}_n$; then $\mathcal{G}$ also is a filter on X.

(c) The formula for $\mathcal{G}_{n+1}$ ensures that $\lim_{g\to\mathcal{G}_{n+1}}(g(x_n)|z_n) = \alpha_n$; it follows that $\lim_{g\to\mathcal{G}}(g(x_n)|z_n) = \alpha_n$ for every n, and $\lim_{g\to\mathcal{G}}(g(x)|z)$ is defined for every $x \in D_1$ and $z \in D_2$.

(d) It follows that $\lim_{g\to\mathcal{G}}(g(x)|z)$ is defined for every $x \in H_1$ and $z \in H_2$. To see this, I appeal to Proposition 4F. Take any $\epsilon > 0$. Set

$$\delta = \min(1, \tfrac{\epsilon}{\|f(a)\|+\gamma(\|x-a\|+\|z\|+1)}) > 0.$$

Then there are $x' \in D_1$, $z' \in D_2$ such that $\|x - x'\| \le \delta$ and $\|z - z'\| \le \delta$. In this case, for any $g \in F_{\{x,x'\}}$,

$$\begin{aligned}
|(g(x)|z) - (g(x')|z')| &\le |(g(x)|z) - (g(x)|z')| + |(g(x)|z') - (g(x')|z')| \\
&= |(g(x)|z - z')| + |(g(x) - g(x')|z')| \\
&\le \|g(x)\|\|z - z'\| + \|g(x) - g(x')\|\|z'\| \\
&\le (\|f(a)\| + \gamma\|x - a\|)\delta + \gamma\|x - x'\|(\|z\| + \|z' - z\|) \\
&\le (\|f(a)\| + \gamma\|x - a\|)\delta + \gamma\delta(\|z\| + 1) \le \epsilon.
\end{aligned}$$

But $F_{\{x,x'\}} \in \mathcal{G}$ and $\lim_{g\to\mathcal{G}}(g(x')|z')$ is defined. Thus the condition of 4F is satisfied and we have a limit $\lim_{g\to\mathcal{G}}(g(x)|z)$.

(e) We have reached the same position as at the beginning of part (e) of the proof in 3F, and can find an extension of f as before.

(f) This proves the theorem when H_2 is separable. For the general case, I use Lemma 4G, as follows. First, by 4Gb, we have an extension of f to a function $f_1 : \overline{A} \to H_2$ such that $\|f_1(x) - f_1(x')\| \le \gamma\|x - x'\|$ for all x, $x' \in \overline{A}$. Next, by 4Ga, $\overline{A}$ has a countable dense subset D_0 say. In this case, the set D_2 of rational linear combinations of elements of $f[D_0]$ is countable, and $\overline{D}_2$ is a closed linear subspace of H_2, so is a Hilbert space in its own right. Now (a)-(e) tell us that there is a suitable extension of f_1 to a function from H_1 to $\overline{D}_2 \subseteq H_2$.

4I Remarks (a) The method in 4H will work whenever H_1 is provided with a 'well-orderable' dense set D_1, rather than just a countable one; here I say that D is well-orderable if there is a well-ordering on D, that is, if there is an ordinal equipollent with D. The point is that (having well-ordered D_1) we can use this in the argument of 1Ib to get a well-ordered dense subset D_0 of $\overline{A}$, from which we can build well-orderings of D_2 and $D_1 \times D_2$. The last of these will give us an order in which to take pairs (x, z) in an inductive construction of $\mathcal{G}$, as in part (b) of the proof above.

(b) Note that the argument there includes a useful general fact:

If X is a set, $\mathcal{G}_0$ is a filter on X, and $\phi : X \to \mathbb{R}$ is a function such that $\limsup_{t\to\mathcal{G}_0}\phi(t)$ is finite, then there is a filter $\mathcal{G}_1$ on X, including $\mathcal{G}_0$, such that $\lim_{t\to\mathcal{G}_1}\phi(t)$ is defined in $\mathbb{R}$.

But it would have done us no good to write this out as a preparatory lemma. For it is not enough, in 4H, just to know that we shall always be able to extend a filter $\mathcal{G}_n$ suitably. We have to have a definite recipe for the extension, so that we nowhere have to make infinitely many choices.

5 Isometries

Kirszbraun's theorem is about Lipschitz functions. It is generally stated in terms of an arbitrary Lipschitz constant $\gamma \geq 0$, but it is very easy to see that it is enough to consider the case $\gamma = 1$, as indeed I do in the foundations 1B and 1D of the proofs above. If however we have the very much stronger hypothesis

$$\|f(x) - f(y)\| = \|x - y\| \text{ for all } x,\, y \in A,$$

that is, f is an **isometry**, in place of '$\|f(x) - f(y)\| \leq \|x - y\|$ for all x, $y \in A$', we get a correspondingly more powerful extension theorem (5D below), which I do not find wholly obvious. So I set it out here.

5A Lemma Let H be a Hilbert space, x, y, z points of H, and $\alpha \in \mathbb{R}$. Then

$$\|z - ((1-\alpha)x + \alpha y)\|^2 = (\alpha^2 - \alpha)\|x - y\|^2 + (1-\alpha)\|z - x\|^2 + \alpha\|z - y\|^2.$$

proof

$$\begin{aligned}
\|z - (1-\alpha)x - \alpha y\|^2 &= \|(1-\alpha)(z - x) + \alpha(z - y)\|^2 \\
&= (1-\alpha)^2\|z - x\|^2 + \alpha^2\|z - y\|^2 + 2\alpha(1-\alpha)(z - x|z - y) \\
&= (1-\alpha)^2\|z - x\|^2 + \alpha^2\|z - y\|^2 \\
&\qquad + \alpha(1-\alpha)(\|z - x\|^2 + \|z - y\|^2 - \|(z - x) - (z - y)\|^2) \\
&= (1-\alpha)^2\|z - x\|^2 + \alpha^2\|z - y\|^2 \\
&\qquad + \alpha(1-\alpha)(\|z - x\|^2 + \|z - y\|^2 - \|x - y\|^2) \\
&= (1-\alpha)\|z - x\|^2 + \alpha\|z - y\|^2 + (\alpha^2 - \alpha)\|x - y\|^2.
\end{aligned}$$

Remark The actual formula here is of no significance. All that matters is that $\|z - ((1-\alpha)x + \alpha y)\|$ can be calculated from α, $\|x - y\|$, $\|z - x\|$ and $\|z - y\|$; and this is because in Hilbert space, as in two-dimensional Euclidean space, triangles are determined up to congruence by the lengths of their sides.

5B Lemma Let H_1 and H_2 be Hilbert spaces, $A \subseteq H_1$ a set and $f : A \to H_2$ an isometry. Suppose that x, $y \in A$ and $\alpha \in \mathbb{R}$ are such that $z = (1-\alpha)x + \alpha y$ belongs to A. Then $f(z) = (1-\alpha)f(x) + \alpha f(y)$.

proof

$$\|f(z) - ((1-\alpha)f(x) + \alpha f(y))\|^2$$
$$= (\alpha^2 - \alpha)\|f(x) - f(y)\|^2 + (1-\alpha)\|f(z) - f(x)\|^2 + \alpha\|f(z) - f(y)\|^2$$

(by Lemma 5A)

$$= (\alpha^2 - \alpha)\|x - y\|^2 + (1-\alpha)\|z - x\|^2 + \alpha\|z - y\|^2$$

(because f is an isometry)

$$= \|z - ((1-\alpha)x + \alpha y)\|^2$$

(by Lemma 5A again, or otherwise)

$$= 0.$$

So $f(z) - ((1-\alpha)f(x) + \alpha f(y)) = 0$ and $f(z) = (1-\alpha)f(x) + \alpha f(y)$.

5C Lemma Let H_1 and H_2 be Hilbert spaces, $A \subseteq H_1$ a set and $f : A \to H_2$ an isometry. Suppose that $w \in H_1$ is expressible as $(1-\alpha)x + \alpha y$ where x, $y \in A$ and $\alpha \in \mathbb{R}$. Then there is a unique isometry $g : A \cup \{w\} \to H_2$ extending f.

proof If $w \in A$ then we must take $g = f$. Otherwise, we can define $g : A \cup \{w\} \to H_2$ by setting $g(z) = f(z)$ for $z \in A$ and $g(w) = (1-\alpha)f(x) + \alpha f(y)$. If $z \in A$, then

$$\begin{aligned}
\|g(z) - g(w)\|^2 &= \|f(z) - ((1-\alpha)f(x) + \alpha f(y))\|^2 \\
&= (\alpha^2 - \alpha)|f(x) - f(y)\|^2 + (1-\alpha)\|f(z) - f(x)\|^2 + \alpha\|f(z) - f(y)\|^2
\end{aligned}$$

(by Lemma 5A)

$$= (\alpha^2 - \alpha)|x) - y\|^2 + (1-\alpha)\|z - x\|^2 + \alpha\|z - y\|^2 = \|z - w\|^2$$

so $\|g(z) - g(w)\| = \|z - w\|$. Now if $z,\ z' \in A \cup \{w\}$,

— either $z,\ z' \in A$ and $\|g(z) - g(z')\| = \|f(z) - f(z')\| = \|z - z'\|$,

— or $z \in A$ and $z' = w$ and $\|g(z) - g(z')\| = \|g(z) - g(w)\| = \|z - w\| = \|z - z'\|$,

— or $z = w$ and $z' \in A$ and $\|g(z) - g(z')\| = \|g(z') - g(w)\| = \|z' - w\| = \|z - z'\|$,

— or $z = z' = w$ and $\|g(z) - g(z')\| = 0 = \|z - z'\|$.

Thus g is an isometry. To see that it is unique, let $h : A \cup \{w\} \to H_2$ be another isometry extending f; then Lemma 5B tells us that

$$h(w) = (1-\alpha)h(x) + \alpha h(y) = (1-\alpha)f(x) + \alpha f(y) = g(w),$$

so $h = g$.

5D Theorem Let H_1 and H_2 be Hilbert spaces, $A \subseteq H_1$ a set and $f : A \to H_2$ an isometry. Let K be the closed affine subspace of H_1 generated by A. Then f has a unique extension to an isometry from K to H_2.

proof (a) I will say that a string $(z_0, \dots, z_n)$ in H_1 is a **determining chain** if for every $i \le n$ there are $x,\ y \in A \cup \{z_j : j < i\}$ and $\alpha \in \mathbb{R}$ such that $z_i = (1-\alpha)x + \alpha y$. In this case we have a unique isometry $g : A \cup \{z_i : i \le n\} \to H_2$ extending f. To see this, induce on n. If $n = 0$, then z_0 is expressible as $(1-\alpha)x + \alpha y$ where $x,\ y \in A$. By Lemma 5C, there is a unique isometry from $A \cup \{z_0\}$ to H_2 extending f. For the inductive step to $n > 0$, the inductive hypothesis tells us that there is a unique isometry $g_0 : A \cup \{z_i : i < n\} \to H_2$ extending f, and now we can apply Lemma 5C to g_0 to see that there is a unique isometry $g : A \cup \{z_i : i \le n\} \to H_2$ extending g_0. Now if $h : A \cup \{z_i : i \le n\} \to H_2$ is an isometry extending f, $h{\restriction}A \cup \{z_i : i < n\}$ is also an isometry extending f, so is equal to g_0, and h extends g_0, so must be equal to g. Thus g is the only isometry from $A \cup \{z_i : i \le n\}$ to H_2 extending f, and the induction continues.

(b) Write K_0 for the set of those $z \in H_1$ such that there is a determining chain $(z_0, \dots, z_n)$ with $z_n = z$. Observe that $K_0 \supseteq A$, because if $z \in A$ then (z) is a determining chain. If $z,\ z' \in K_0$ and $\alpha \in \mathbb{R}$ then $w = (1-\alpha)z + \alpha z'$ belongs to K_0. For we have determining chains $(z_0, \dots, z_m)$ and $(z'_0, \dots, z'_n)$ such that $z_m = z$ and $z'_n = z'$, and now $(z_0, \dots, z_m, z'_0, \dots, z'_n, w)$ is a determining chain, so $w \in K_0$.

(c) If $z \in K_0$, $(z_0, \dots, z_m)$ and $(z'_0, \dots, z'_n)$ are determining chains such that $z_m = z'_n = z$, and $g : A \cup \{z_i : i \le m\} \to H_2$, $g' : A \cup \{z'_i : i \le n\} \to H_2$ are isometries extending f, then $g(z) = g'_z$. For, just as in (b), $(z_0, \dots, z_m, z'_0, \dots, z'_n)$ is a determining chain, so there is an isometry $\tilde{g} : A \cup \{z_0, \dots, z_m, z'_0, \dots, z'_n\} \to H_2$ extending f; we must have $\tilde{g}{\restriction}A \cup \{z_0, \dots, z_m\} = g$ and $\tilde{g}{\restriction}A \cup \{z'_0, \dots, z'_n\} = g'$, so

$$g(z) = \tilde{g}(z_m) = \tilde{g}(z) = \tilde{g}(z'_n) = g'(z).$$

(d) We can therefore define a function $h : K_0 \to H_2$ by saying that $h(z) = g(z_n)$ whenever $(z_0, \dots, z_n)$ is a determining chain with $z_n = z$ and $g : A \cup \{z_0, \dots, z_n\} \to H_2$ is an isometry. Now h is an isometry. For if $z,\ z' \in K_0$ and $(z_0, \dots, z_m)$, $(z'_0, \dots, z'_n)$ are determining chains with $z = z_m$ and $z' = z'_n$, $(z_0, \dots, z_m, z'_0, \dots, z'_n)$ is a determining chain, so $h{\restriction}A \cup \{z_0, \dots, z_m, z'_0, \dots, z'_n\}$ is an isometry, and

$$\|h(z) - h(z')\| = \|h(z_m) - h(z'_n)\| = \|z_m - z'_n\| = \|z - z'\|.$$

(e) In fact h is the only isometry from $K_0 \to H_2$ extending f. For if $h' : K_0 \to H_2$ is an isometry from K_0 to H_2 and $z \in K_0$, there is a determining chain $(z_0, \dots, z_n)$ with $z = z_n$, and now $z_i \in K_0$ for every $i \le n$, so $h{\restriction}A \cup \{z_0, \dots, z_n\}$ and $h'{\restriction}A \cup \{z_0, \dots, z_n\}$ are both isometries extending f; by (a) they are equal and $h'(z) = h'(z_n) = h(z_n) = h(z)$.

(f) Being an isometry, h is surely uniformly continuous, while H_2 is complete. So h has an extension to a uniformly continuous function $\tilde{h}$ from the norm-closure K of K_0 to H_2. The set $\{(z, z') : z,\ z' \in K, \|\tilde{h}(z) - \tilde{h}(z')\| = \|z - z'\|\}$ is closed in $K \times K$ and includes the dense subset $K_0 \times K_0$ so is the whole of $K \times K$, and $\tilde{h}$ is an isometry.

Similarly, if $\alpha \in \mathbb{R}$, the set $\{(z, z') : z,\ z' \in K,\ (1-\alpha)z + \alpha z' \in K\}$ is closed in $K \times K$ and includes $K_0 \times K_0$ (by (b) above), so is the whole of $K \times K$. This shows that K is an affine subspace of H_1, and of course it is closed and includes A.

(g) If K' is any other closed affine subspace of H_1 including A, then K' must include K_0, so $K' \supseteq K$. Thus K is the smallest closed affine subspace of H_1 including A, that is, it is the closed affine subspace of H_1 generated by A.

(h) Finally, if $\tilde{h}' : K \to H_2$ is any other isometry extending f, $\tilde{h}' \restriction K_0$ extends f, so is equal to h. Now $\{z : z \in K, \tilde{h}'(z) = \tilde{h}(z)\}$ is a closed subset of K (because $\tilde{h}'$ and $\tilde{h}$ are both continuous) and includes K_0 (by (e)), so is equal to K, and $\tilde{h}' = \tilde{h}$. Thus $\tilde{h}$ is the only isometry extending f to K, and the theorem is proved.

5E Remarks Most of the work in 5A-5D is just algebra of inner product spaces. The point is that the affine operation $(x, y, \alpha) \mapsto (1-\alpha)x + \alpha y$ is determined by the metric structure; $z = (1-\alpha)x + \alpha y$ iff $(\alpha^2 - \alpha)\|x - y\|^2 + (1-\alpha)\|z - x\|^2 + \alpha\|z - y\|^2 = 0$.

I have spelt the proof of 5D out carefully, with 'determining chains', to show that it does not depend on any form of the axiom of choice. When we come to the final stage (parts (f)-(h) of the proof of 5D) we ought of course to check that our favourite theorems about uniform continuity don't depend on being able to choose sequences. This will mean that, as in 4B, we need to use the Cauchy-filter definition of completeness in H_2 (it doesn't matter whether H_1 is complete or not). Actually the statement of the theorem, claiming that the extension is unique, is a strong hint that there ought to be a proof which does not depend on magical assistance; if any jinn would lead us to the same place, we shouldn't need their help.

Acknowledgement Conversation with B.Kirchheim; correspondence with M.Morillon.

References

Federer H. [69] *Geometric Measure Theory.* Springer, 1969 (reprinted 1996).

Halpern J.D. & Levy A. [71] 'The Boolean prime ideal theorem does not imply the axiom of choice', pp. 83-134 in Scott 71.

Kirszbraun M.D. [34] 'Über die zusammenziehenden und Lipschitzian Transformationen', Fundamenta Math. 22 (1934) 77-108 (`http://matwbn.icm.edu.pl/ksiazki/fm/fm22/fm22112.pdf`).

Scott D.S. [71] (ed.) *Axiomatic Set Theory.* Amer. Math. Soc., 1971 (Proceedings of Symposia in Pure Mathematics XIII, vol. 1).

Jech T. [73] *The Axiom of Choice.* North-Holland, 1973.