UNIFORM EXTENSIONS

MARK H. KIM

Abstract. This is a survey of a few uniform and Lipschitz extension results, based on the first few weeks of Assaf Naor's Spring 2014 lectures on geometric nonlinear functional analysis at the Courant Institute. As I have made modifications, rearrangements, and additions as I saw fit, the present notes do not necessarily reflect Naor's viewpoints on the material or reproduce the lectures accurately; in particular, all the errors herein are wholly my own.

Contents

1. Extension of real-valued functions

We begin by introducing a quantitative measure of different notions of continuity we encounter in mathematics.

Definition 1.1. Let (X, d_X) and (Y, d_Y) be metric spaces. The *modulus of continuity* of a continuous function $f : X \to Y$ is the function

$$\omega_f(t) = \sup\{d_Y(f(x), f(y)) : x, y \in X \text{ and } d_X(x, y) \leq t\}.$$

Recall that $f : (X, d_X) \to (Y, d_Y)$ is *uniformly continuous* if, for each $\varepsilon > 0$, we can find a $\delta > 0$ such that $x, y \in X$ and $d_X(x, y) \leq \delta$ imply that $d_Y(f(x), f(y)) \leq \varepsilon$. It immediately follows that $\omega_f(\delta) \leq \varepsilon$, whence ω_f is bounded on some subinterval $[0, t_0)$. Moreover, by taking ε to be arbitrarily small, we see that $\lim_{t\to 0^+} \omega_f(t) = 0$. We remark that this is the defining property of uniformly continuous functions. Indeed, if $\lim_{t\to 0^+} \omega_f(t) = 0$, then, for each $\varepsilon > 0$, we can find $\delta > 0$ such that $\omega_f(t) \leq \varepsilon$ for all $t \in [0, \delta]$. This, in particular, implies that $d_Y(f(x), f(y)) \leq \varepsilon$ whenever $d_X(x, y) \leq \delta$.

Recall also that f is *α-Hölder continuous* for some $0 < \alpha \leq 1$ if there exists a constant k such that $d_Y(f(x), f(y)) \leq k d_X(x, y)^\alpha$ for all $x, y \in X$. It follows at once that f is α-Hölder continuous if and only if there exists a constant k such that $\omega_f(t) \leq kt^\alpha$ for all $t \in [0, \infty)$. Setting $\alpha = 1$, we obtain a quantitative characterization of *Lipschitz continuity*: f is Lipschitz continuous if and only if there exists a *Lipschitz constant* k such that $\omega_f(t) \leq kt$ for all $t \in [0, \infty)$.

Existence of dominating functions can be used to characterize other notions of continuity. For example, a collection of functions $\{f_\alpha : (X, d_X) \to (Y, d_Y)\}_\alpha$ admits a non-increasing function $\omega : [0, \infty) \to [0, \infty]$ such that $\lim_{t \to 0^+} \omega(t) = 0$ and $\omega_{f_\alpha}(t) \leq \omega(t)$ for all α if and only if $\{f_\alpha\}_\alpha$ is *equicontinuous*, viz., each $\varepsilon > 0$ admits a $\delta > 0$ such that $x, y \in X$ and $d_X(x, y) \leq \delta$ imply $d_Y(f_\alpha(x), f_\alpha(y)) \leq \varepsilon$ for all α.

We also note that the dominating functions of Hölder-continuous functions we have considered above are *subadditive*, i.e., $\omega(t_1 + t_2) \leq \omega(t_1) + \omega(t_2)$. To shed light on the significance of this condition, we recall the Hahn-Banach theorem, which concerns the existence of extensions of linear functionals dominated by sublinear functionals[1]. Integral to this extension result is subadditivity of dominating functions, as the following nonlinear generalization shows.

Theorem 1.2 (Nonlinear Hahn-Banach theorem)**.** *Let (X, d_X) be a metric space and A a subset of X. If a uniformly continuous function $f : A \to \mathbb{R}$ admits a nondecreasing subadditive function $\omega : [0, \infty) \to [0, \infty)$ such that $\lim_{t \to 0^+} \omega(t) = 0$ and $\omega_f \leq \omega$, then there exists a uniformly continuous extension $F : X \to \mathbb{R}$ of f such that $\omega_F \leq \omega$.*

Proof. We mimic the proof of the classical Hahn-Banach theorem, which involves a standard Zorn's lemma argument. Indeed, we shall construct a uniform extension of $f : A \to \mathbb{R}$ on $A \cup \{x_0\}$ for some $x_0 \in X \smallsetminus A$, at which point Zorn's lemma will produce a uniform extension on the whole space X.

We therefore suppose that A is a proper subset of X and fix $x_0 \in X \smallsetminus A$. Note that any extension $\tilde{f} : A \cup \{x_0\} \to \mathbb{R}$ of f satisfies the following a priori estimate:

$$\begin{aligned} \omega_{\tilde{f}}(t) &= \max\left(\omega_f(t), \sup\{|\tilde{f}(x_0) - f(x)| : x \in A \text{ and } d_X(x, x_0) \leq t\}\right) \\ &\leq \max\left(\omega(t), \sup\{|\tilde{f}(x_0) - f(x)| : x \in A \text{ and } d_X(x, x_0) \leq t\}\right). \end{aligned}$$

Therefore, if

$$|\tilde{f}(x_0) - f(x)| \leq \omega(d_X(x, x_0)) \tag{1.3}$$

for all $x \in X$, then

$$|\tilde{f}(x_0) - f(x)| \leq \omega(d_X(x, x_0)) \leq \omega(t)$$

whenever $d_X(x, x_0) \leq t$, whence the a priori estimate on $\tilde{f}$ implies that $\omega_{\tilde{f}}(t) \leq \omega(t)$. Since t was arbitrary, we conclude that $\omega_{\tilde{f}} \leq \omega$.

It therefore suffices to find an extension $\tilde{f}$ satisfying (1.3), which we can rewrite as

$$f(x) - \omega(d_X(x, x_0)) \leq \tilde{f}(x_0) \leq f(x) + \omega(d_X(x, x_0)).$$

To this end, we let

$$\alpha = \sup_{x \in A} f(x) - \omega(d_X(x, x_0)) \quad \text{and} \quad \beta = \inf_{x \in A} f(x) + \omega(d_X(x, x_0)). \tag{1.4}$$

[1] We recall that a *sublinear functional* on an $\mathbb{F}$-vector space is a subadditive function $\rho : X \to \mathbb{F}$ such that $\rho(\lambda x) = \lambda \rho(x)$ for all positive scalars λ.

We claim that $\alpha \leq \beta$. Assuming the claim, we can find a γ such that $\alpha \leq \gamma \leq \beta$, whence setting

$$\tilde{f}(x_0) = \begin{cases} f(x) & \text{if } x \in A; \\ \gamma & \text{if } x = x_0 \end{cases}$$

yields a desired extension. To verify the claim, we observe that the subadditivity of ω furnishes the estimate

$$|f(x) - f(y)| \leq \omega_f(d_X(x,y)) \leq \omega(d_X(x,y)) \leq \omega(d_X(x,x_0)) + \omega(d_X(y,x_0))$$

for all $x, y \in A$, whence

$$f(x) - \omega(d_X(x,x_0)) \leq f(y) - \omega(d_X(y,x_0)).$$

Taking the supremum over x and the infimum over y, respectively, we conclude that $\alpha \leq \beta$, as was to be shown. □

Remark. The existence of the dominating function ω is essential. As a counterexample, we consider $f(x) = x^2$ defined on the subset

$$A = \bigcup_{n=1}^{\infty} \left(n, \sqrt{n^2 + \frac{1}{n}}\right)$$

of the real line. By considering f on

$$A_N = \bigcup_{n=1}^{N-1} \left(n, \sqrt{n^2 + \frac{1}{n}}\right)$$

and $A \smallsetminus A_N$ separately, we see that $f : A \to \mathbb{R}$ is uniformly continuous. Note that

$$\omega_f(t) = \begin{cases} t^2 & \text{if } t < 1; \\ \infty & \text{if } t \geq 1; \end{cases}$$

and so ω_f cannot be dominated by a nondecreasing subadditive function ω such that $\lim_{t \to 0^+} \omega(t) = 0$.

We now show that f does not admit a uniform extension on $\mathbb{R}$ with the usual Euclidean metric. To see this, we let F be a continuous extension of f on $\mathbb{R}$. Since $\omega_F \geq \omega_f$, we see that $\omega_F(t) = \infty$ for all $t \in [1, \infty)$. If $0 < t < 1$, then we can find an integer N such that $t > 1/N$. For each $n \in \mathbb{N}$, $F(n) = n^2$ and $F(n+1) = n^2$, whence the continuous extension F should increase at least by $[(n+1)^2 - n^2]/N$ on one of the intervals $\{[n+k/N, n+(k+1)/N] : 0 \leq k \leq n-1\}$. Since $\lim_{n \to \infty}[(n+1)^2 - n^2]/N = \infty$, we conclude that $\omega(t) = \infty$. It follows that F is not uniformly continuous. □

Remark. We can avoid the use of Zorn's lemma by taking (1.4) as the definition of our extension: say,

$$F(x) = \begin{cases} f(x) & \text{if } x \in A; \\ \sup_{y \in A} f(y) - \omega(d_X(y,x)) & \text{if } x \notin A. \end{cases}$$

serves as a desired extension. Contrast this with the usual proof of the classical Hahn-Banach theorem, which involves fixing a non-zero vector from each one-dimensional subspace of the full space. Indeed, the classical Hahn-Banach theorem, combined with an extended version of the Krein-Milman theorem, is equivalent to the axiom of choice: see [BJ71]. □

We now specialize the nonlinear Hahn-Banach theorem to the class of real-valued Lipschitz functions to obtain another extension result.

Corollary 1.5. *Let (X, d_X) be a metric space and A a subset of X. If $f : A \to \mathbb{R}$ is Lipschitz, then there exists a Lipschitz extension $F : X \to \mathbb{R}$ of f such that $\|F\|_{\mathrm{Lip}} = \|f\|_{\mathrm{Lip}}$.*

Proof. We set $\omega(t) = \|f\|_{\mathrm{Lip}} t$ and apply Theorem 1.2. to obtain a uniform extension F of f such that $\omega_F \leq \omega$. It immediately follows from the modulus of continuity characterization of Lipschitz continuity that F is Lipschitz, and that $\|F\|_{\mathrm{Lip}} \leq \|f\|_{\mathrm{Lip}}$. The reverse inequality is trivial. □

The two extension results we have established thus far can be generalized to maps taking values in $l_\infty(\Gamma)$.

Corollary 1.6. *Let (X, d_X) be a metric space, A a subset of X, and Γ a nonempty set. If a uniformly continuous function $f : A \to l_\infty(\Gamma)$ admits a nondecreasing subadditive function $\omega : [0, \infty) \to [0, \infty)$ such that $\lim_{t \to 0^+} \omega(t) = 0$ and $\omega_f \leq \omega$, then there exists a uniformly continuous extension $F : X \to l_\infty(\Gamma)$ of f such that $\omega_F \leq \omega$.*

Proof. For each α, we define f_α by setting $f_\alpha(x) = (f(x))(\alpha)$ for all $x \in A$. Note that

$$\omega_{f_\alpha}(t) = \sup_{\substack{x,y \in A \\ d_X(x,y) \leq t}} |f_\alpha(x) - f_\alpha(y)| \leq \sup_{\substack{x,y \in A \\ d_X(x,y) \leq t}} \|f(x) - f(y)\|_\infty = \omega_f(t)$$

for each $t \in [0, \infty)$, and so $\omega_{f_\alpha} \leq \omega$. We can therefore apply the nonlinear Hahn-Banach theorem (Theorem 1.2) to obtain a uniform extension $F_\alpha : X \to \mathbb{R}$ of f_α such that $\omega_{F_\alpha} \leq \omega$.

We now define $F : X \to l_\infty(\Gamma)$ by setting $(F(x))(\alpha) = F_\alpha(x)$ for each $x \in X$. Whenever $x, y \in X$ and $d_X(x, y) \leq t$, we see that

$$\|F(x) - F(y)\|_\infty = \sup_{\alpha \in \Gamma} |F_\alpha(x) - F_\alpha(y)| \leq \sup_{\alpha \in \Gamma} \omega(t) = \omega(t),$$

whence taking the supremum over all $x, y \in X$ with $d_X(x, y) \leq t$ furnishes the estimate

$$\omega_F(t) \leq \omega(t).$$

It follows that F is a desired uniform extension of f. □

Corollary 1.7. *Let (X, d_X) be a metric space, A a subset of X, and Γ a nonempty set. If $f : A \to l_\infty(\Gamma)$ is Lipschitz, then there exists a Lipschitz extension $F : X \to l_\infty(\Gamma)$ of f such that $\|F\|_{\mathrm{Lip}} = \|f\|_{\mathrm{Lip}}$.*

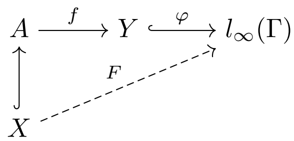
Proof. Proceed as in the proof of Corollary 1.5. □

In a way, Corollary 1.6 shows that every uniformly continuous function $f : (A, d_X) \to (Y, d_Y)$ with a dominating function ω for its modulus of continuity admits a uniform extension on (X, d_X) regardless of its codomain. As we shall see momentarily, every metric space can be embedded isometrically into some $l_\infty(\Gamma)$, whence the composite map

$$A \xrightarrow{\ f\ } Y \xhookrightarrow{\ \varphi\ } l_\infty(\Gamma)$$

is uniformly continuous with the same dominating function ω for its modulus of continuity. Corollary 1.6 can then be applied to the above composite map to obtain a uniform extension F such that $\omega_F \leq \omega$. See the diagram below:

(1.8)



Proposition 1.9. *For each nonempty metric space (X, d_X), we can find a set Γ and an isometry $\varphi : X \to l_\infty(\Gamma)$.*

Proof. We fix $x_0 \in X$, let $\Gamma = X$, and define $(\varphi(x))(y) = d_X(x, y) - d_X(x_0, y)$. For each fixed $x \in X$, the triangle inequality states that $d_X(x, y) \leq d_X(x, x_0) + d_X(x_0, y)$ for all $y \in X$, whence $\|\varphi(x)\|_\infty \leq d_X(x, x_0)$. Moreover,

$$\|\varphi(x_1) - \varphi(x_2)\|_\infty = \sup_{y \in X} d_X(x_1, y) - d_X(x_2, y) = d_X(x_1, x_2),$$

and so φ is an isometry into $l_\infty(\Gamma)$. □

2. Non-extendability Results

Nevertheless, the process of uniform extension carried out in (1.8) does not guarantee that $\operatorname{im} F \subseteq \varphi(Y)$. As such, constant-preserving extensions such as Corollary 1.5 do not occur often. Let us study a constant-increasing Lipschitz extension and its generalizations in detail.

Example 2.1. We refer the reader to Section 5 of the Appendix for terminology. Consider the complete graph K^4 with weights on its edges given as follows:

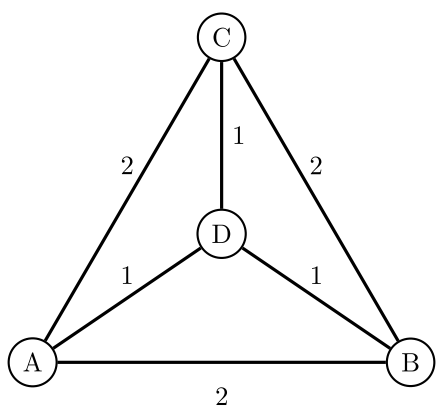


Figure 1. Metrized K^4

The shortest path metric d induced by the above weight structure turns K^4 into a metric space. We now let $M = \{A, B, C\}$ be a subspace of the metric space (K^4, d) and consider the map $f : M \to \mathbb{R}^2$ that sends M to an equilateral triangle $A'B'C'$ of side length 2 in $\mathbb{R}^2$ with the usual Euclidean metric. Evidently, f is an isometry, hence Lipschitz with $\|f\|_{\text{Lip}} = 1$.

Now, we consider a Lipschitz extension $F : K^4 \to \mathbb{R}^2$ of f. We claim that $\|F\|_{\text{Lip}} > 1$. Indeed, trigonometry tells us that the distance from $F(D) = D'$ to any of the three vertices must be at least $2/\sqrt{3}$, whence $\|F\|_{\text{Lip}} \geq \frac{2}{\sqrt{3}}$.

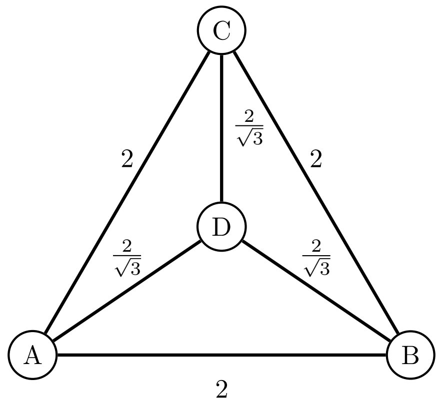


FIGURE 2. A "minimal" configuration of $F(K^4)$

Example 2.2 ([JLS86]). Let us now generalize Example 2.1 to exhibit constant-increasing Lipschitz-extension phenomena on graphs on finite metric spaces. Once again, we refer the reader to Section 5 of the Appendix for graph-theoretic terminology.

Let (X, d) be the N-dimensional Hamming cube $\mathbb{Z}_2^N = \{-1, 1\}^N$ (Appendix, Definition 7.7). Let $(\tilde{X}, \tilde{d})$ be a disjoint copy of (X, d), and, for each $x \in X$, we write $\tilde{x}$ to denote the corresponding copy of x in $\tilde{X}$. We let C and r be two positive constants that will be determined later. Define a weighted graph on $X \cup \tilde{X}$ as follows:

(a) if $x, y \in X$ are distinct, then they are adjacent and the length of the corresponding edge is $\sqrt{d(x, y)}$;
(b) if $\tilde{x}, \tilde{y} \in \tilde{X}$ are distinct, then they are adjacent and the length of the corresponding edge is $\frac{\tilde{d}(\tilde{x}, \tilde{y})}{Cr}$;
(c) each $x \in X$ is adjacent to $\tilde{x} \in X$, and the length of the corresponding edge is r.

We let ρ be the graph metric induced from the weight structure on $X \cup \tilde{X}$. We pick C so that

$$\rho(x, y) = \sqrt{d(x, y)} \tag{2.3}$$

whenever $x, y \in X$. To see that this is possible, we note that every path from x to y that leaves X is of length at least $r + \frac{\tilde{d}(\tilde{x}, \tilde{y})}{Cr} + r$. Therefore, if

$$2r + \frac{d(x, y)}{Cr} = 2r + \frac{\tilde{d}(\tilde{x}, \tilde{y})}{Cr} \geq \sqrt{d(x, y)} \tag{2.4}$$

for all $x, y \in X$, then (2.3) holds. Now, if $C = 8$, then

$$d(x, y) - Cr\sqrt{d(x, y)} + 2Cr^2 = d(x, y) - 8r\sqrt{d(x, y)} + 16r^2 = (\sqrt{d(x, y)} - 4r)^2 \geq 0,$$

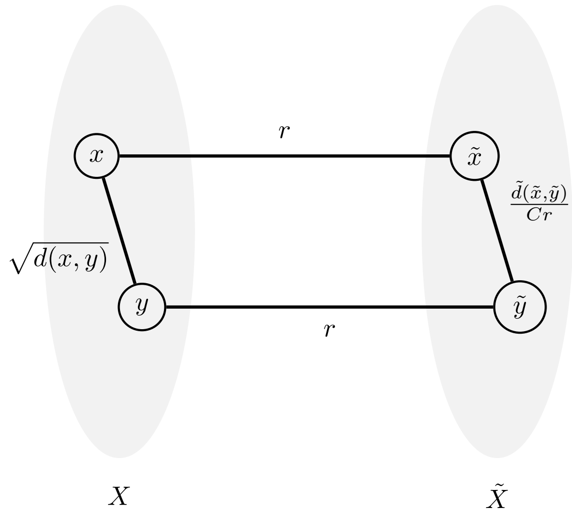


FIGURE 3. Weighted graph structure on $X \cup \tilde{X}$

and so

$$d(x,y) - Cr\sqrt{d(x,y)} + 2Cr^2 \geq 0$$

holds for all $x, y \in X$. This is equivalent to (2.4), and so (2.3) holds for all $x, y \in X$ if $C = 8$.

We now consider the function $f : X \to \mathbb{R}^N$ given by $f(x) = (x_1, \ldots, x_N)$, where $\mathbb{R}^N$ is endowed with the usual Euclidean norm $\|\cdot\|_2$. (2.3) implies that

$$\|f(x) - f(y)\|_2 = \left(\sum_{n=1}^{N} |x_n - y_n|^2\right)^{1/2} \leq \left(\sum_{n=1}^{N} 2|x_n - y_n|\right)^{1/2} = 2\rho(x,y)$$

for all $x, y \in X$, whence $\|f\|_{\mathrm{Lip}} \leq 2$.

Let $F = (F_1, \ldots, F_N)$ be a Lipschitz extension of f on $X \cup \tilde{X}$. We shall show that

$$\frac{2\sqrt{N}}{2r + \sqrt{N}/Cr} \leq \|F\|_{\mathrm{Lip}}. \tag{2.5}$$

If we let $r = N^{1/4}$, then the above estimate becomes

$$\frac{2}{2 + 1/C} N^{1/4} \leq \|F\|_{\mathrm{Lip}},$$

whence $\|F\|_{\mathrm{Lip}}$ can be made as large as desired by choosing a large enough N. In particular, it is possible to choose r and N such that $\|F\|_{\mathrm{Lip}} > 2 = \|f\|_{\mathrm{Lip}}$ for *all* Lipschitz extensions F of f.

To establish (2.5), we use the fact that the Hamming cube is of Enflo type 2 (Appendix, Example 7.12), which, in particular, implies that

$$\sum_{\tilde{x} \in \tilde{X}} (F_k(\tilde{x}) - F_k(-\tilde{x}))^2 \leq \sum_{\tilde{x} \in \tilde{X}} \sum_{n=1}^{N} (F_k(\tilde{x}_1, \ldots, -\tilde{x}_n, \ldots, \tilde{x}_N) - F_k(\tilde{x}))^2.$$

for all $1 \le k \le N$. Therefore, we obtain the estimate

$$\begin{aligned}
\sum_{\tilde{x}\in\tilde{X}} \|F(\tilde{x}) - F(-\tilde{x})\|_2^2 &= \sum_{\tilde{x}\in X}\sum_{k=1}^{N}(F(\tilde{x}) - F(-\tilde{x}))^2 \\
&= \sum_{k=1}^{N}\sum_{\tilde{x}\in X}(F(\tilde{x}) - F(-\tilde{x}))^2 \\
&\le \sum_{k=1}^{N}\sum_{\tilde{x}\in\tilde{X}}\sum_{n=1}^{N}(F_k(\tilde{x}_1,\dots,-\tilde{x}_n,\dots,\tilde{x}_N) - F_k(\tilde{x}))^2 \\
&= \sum_{\tilde{x}\in\tilde{X}}\sum_{n=1}^{N}\sum_{k=1}^{N}(F_k(\tilde{x}_1,\dots,-\tilde{x}_n,\dots,\tilde{x}_N) - F_k(\tilde{x}))^2 \\
&= \sum_{\tilde{x}\in\tilde{X}}\sum_{n=1}^{N}\|F(\tilde{x}_1,\dots,-\tilde{x}_n,\dots,\tilde{x}_N) - F(\tilde{x})\|_2^2.
\end{aligned}$$

We shall obtain (2.5) by computing upper and lower bounds of the estimate

$$\sum_{\tilde{x}\in\tilde{X}} \|F(\tilde{x}) - F(-\tilde{x})\|_2^2 \le \sum_{\tilde{x}\in\tilde{X}}\sum_{n=1}^{N}\|F(\tilde{x}_1,\dots,-\tilde{x}_n,\dots,\tilde{x}_N) - F(\tilde{x})\|_2^2 \tag{2.6}$$

obtained above. Observe that $F|_X = f$, and so

$$\begin{aligned}
\|F(\tilde{x}) - F(-\tilde{x})\|_2 &= \|F(\tilde{x}) - F(x) + F(x) - F(-x) + F(-x) - F(-\tilde{x})\|_2 \\
&= \|[F(x) - F(-x)] - [F(x) - F(\tilde{x})] - [F(-\tilde{x}) - F(-x)]\|_2 \\
&\ge \|F(x) - F(-x)\|_2 - \|F(x) - F(\tilde{x})\|_2 - \|F(-\tilde{x}) - F(-x)\|_2 \\
&\ge \|F(x) - F(-x)\|_2 - \|F\|_{\mathrm{Lip}}\,\rho(x,\tilde{x}) - \|F\|_{\mathrm{Lip}}\,\rho(-\tilde{x},-x) \\
&= \|f(x) - f(-x)\|_2 - 2r\|F\|_{\mathrm{Lip}}.
\end{aligned}$$

Since

$$\|f(x) - f(-x)\|_2^2 = \sum_{k=1}^{N}(f_k(x) - f_k(-x))^2 = \sum_{k=1}^{N} 4x_k^2 = 4N,$$

we see that

$$\|F(\tilde{x}) - F(-\tilde{x})\|_2 \ge 2\sqrt{N} - 2r\|F\|_{\mathrm{Lip}}. \tag{2.7}$$

We also note that

$$\begin{aligned}
\|F(\tilde{x}_1,\dots,-\tilde{x}_n,\dots,\tilde{x}_N) - F(\tilde{x})\|_2^2 &\le \|F\|_{\mathrm{Lip}}\,\rho\left((\tilde{x}_1,\dots,-\tilde{x}_n,\dots,\tilde{x}_N),\tilde{x}\right) \\
&= \frac{\|F\|_{\mathrm{Lip}}}{Cr}\tilde{d}\left((\tilde{x}_1,\dots,-\tilde{x}_n,\dots,\tilde{x}_N),\tilde{x}\right) \\
&= \frac{\|F\|_{\mathrm{Lip}}}{Cr}.
\end{aligned}$$

By applying above estimate and (2.7) to (2.6), we obtain the estimate

$$\sum_{\tilde{x}\in X}\left(2\sqrt{N} - 2r\|F\|_{\mathrm{Lip}}\right)^2 \le \sum_{\tilde{x}\in X}\sum_{n=1}^{N}\left(\frac{\|F\|_{\mathrm{Lip}}}{Cr}\right)^2.$$

Writing out the sums, we can rewrite the above estimate as

$$2^N \left(2\sqrt{N} - 2r\|F\|_{\mathrm{Lip}}\right)^2 \le N2^N \left(\frac{\|F\|_{\mathrm{Lip}}}{Cr}\right)^2,$$

and a simple rearrangement now yields (2.5). □

Example 2.8. Modifying Example 2.2, we can also show that Hilbert-valued α-Hölder functions do not always admit α-Hölder extensions. The following counterexample is valid for $1/2 < \alpha \le 1$; we will show in Corollary 3.11 that Hilbert-valued α-Hölder functions are always extendible if $0 < \alpha \le 1/2$. □

3. Extension of Hilbert-valued functions

The above examples show that uniform extensions in general are not constant-preserving, even if they are Hilbert-valued. If, however, we restrict the class of domains, we see that all Hilbert-valued Lipschitz functions admit Lipschitz extensions.

Theorem 3.1 (Kirszbraun extension theorem, 1934; [Kir34])**.** *Let H_1 and H_2 be a Hilbert space and A a subset of H_1. If $f : A \to H_2$ is Lipschitz, then there exists a Lipschitz extension $F : H_1 \to H_2$ of f such that $\|F\|_{\mathrm{Lip}} = \|f\|_{\mathrm{Lip}}$.*

To prove the Kirszbraun extension theorem, it suffices to extend f by one point, similarly as in the proof of the nonlinear Hahn-Banach theorem (Theorem 1.2). Zorn's lemma then takes care of the rest.

We therefore suppose that A is a proper subset of H_1 and let $x_* \in H_1 \smallsetminus A$. By considering $\|f\|_{\mathrm{Lip}}^{-1} f$ if necessary, we can assume without loss of generality that $\|f\|_{\mathrm{Lip}} = 1$.

Consider the collection of balls $\{B_{H_2}(f(x); \|x - x_*\|)\}_{x \in A}$. If the intersection of this collection is nonempty, then the extension $F : A \cup \{x_*\} \to H_2$ of f defined by setting $F(x_*)$ to be an arbitrary point in the intersection $\cap_{x\in A} B_{H_2}(f(x); \|x - x_*\|)$ is 1-Lipschitz. Indeed, $F(x_*) \in B_{H_2}(f(x); \|x - x_*\|)$ for all $x \in A$, and so

$$\|F(x) - F(x_*)\| = \|f(x) - F(x_*)\| \le \|x - x_*\|,$$

for all $x \in A$.

It thus suffices to show that

$$\cap_{x\in A}\, B_{H_2}(f(x); \|x - x_*\|) \ne \varnothing. \tag{3.2}$$

f is 1-Lipschitz, and $x_* \in \cap_{X\in A} B_{H_1}(x; \|x - x_0\|)$, whence (3.2) is a consequence of the following lemma:

Lemma 3.3. *Let H_1 and H_2 be Hilbert spaces, I an index set, $\{x_\beta\}_{\beta\in I}$ a subset of H_1, $\{y_\beta\}_{\beta\in I}$ a subset of H_2, and $\{r_\beta\}_{\beta\in I}$ a collection of nonnegative real numbers. If*

$$\|y_\beta - y_\gamma\| \le \|x_\beta - x_\gamma\| \tag{3.4}$$

for all $\beta, \gamma \in I$, then $\bigcap_{\beta\in I} B_{H_2}(y_\beta, r_\beta) \ne \varnothing$ whenever $\bigcap_{\beta\in I} B_{H_1}(x_\beta, r_\beta) \ne \varnothing$.

Before we prove the lemma, we show that the lemma is, in fact, equivalent to the Kirszbraun extension theorem. Given the collections $\{x_\beta\}_{\beta\in I}$, $\{y_\beta\}_{\beta\in I}$, $\{r_\beta\}_{\beta\in I}$ as above, we define a function $f : \{x_\beta\}_{\beta\in I} \to H_2$ by setting $f(x_\beta) = y_\beta$. (3.4) implies that f is 1-Lipschitz, whence the Kirszbraun extension theorem furnishes a

1-Lipschitz extension $F : H_1 \to H_2$ of f. In particular, if $x_* \in \bigcap_{\beta\in I} B_{H_1}(x_\beta, r_\beta)$, then

$$\|y_\beta - F(x_*)\| = \|F(x_\beta) - F(x_*)\| \leq \|x_\beta - x_*\|$$

for all $\beta \in I$, whence $F(x_*) \in \bigcap_{\beta\in I} B_{H_2}(y_\beta, r_\beta)$. The above lemma is therefore a consequence of the Kirszbraun extension theorem.

Proof of Lemma 3.3. We proceed in three steps.

Step 1. The lemma holds whenever $\dim H_1 < \infty$***,*** $\dim H_2 < \infty$***, and*** $|I| < \infty$.

Let us assume for now that $\dim H_1 < \infty$, $\dim H_2 < \infty$, and $I = \{1, \ldots, N\}$. If $x_k \in \bigcap_{n=1}^N B_{H_1}(x_n, r_n)$, then Hypothesis (3.4) implies that

$$\|y_k - y_n\| \leq \|x_k - x_n\| \leq r_n$$

for all $1 \leq n \leq N$, whence $y_k \in \bigcap_{n=1}^N B_{H_2}(y_n, r_n)$. We therefore suppose that

$$\{x_1, \ldots, x_N\} \cap \left(\bigcap_{n=1}^N B_{H_1}(x_n, r_n) \right) = \varnothing.$$

We fix $x_0 \in \bigcap_{n=1}^N B_{H_1}(x_n, r_n)$, so that $\|x_0 - x_n\| \leq r_n$ for all $1 \leq n \leq N$. If there exists a point $y_0 \in H_2$ such that

$$\|y_0 - y_n\| \leq \|x_0 - x_n\| \tag{3.5}$$

for all $1 \leq n \leq N$, then $y_0 \in \bigcap_{n=1}^N B_{H_2}(y_n, r_n)$. This, in particular, implies that the intersection is nonempty.

To find such a point, we consider the function

$$g(y) = \max_{1\leq n\leq N} \frac{\|y - y_n\|_{H_2}}{\|x_0 - x_n\|_{H_1}}.$$

If we can find a point $y_0 \in H_2$ such that $g(y_0) \leq 1$, then (3.5) holds for all $1 \leq n \leq N$ with the same y_0). It therefore suffices to study the minimum of g, whose existence is established by the claim below.

Claim. *There exists a point $y_0 \in H_1$ such that $g(y_0)$ the global minimum of g.*

Proof of claim. We note that g is continuous. Moreover,

$$\lim_{\|y\|\to\infty} g(y) \geq \lim_{\|y\|\to\infty} \frac{\|y - y_1\|}{\max_{1\leq n\leq N} \|x_0 - x_n\|} = \infty,$$

whence we can find $R, M > 0$ such that

- $g(y) \leq M$ for at least one $y \in H_2$, and
- $g(y) > M$ for all $\|y\| > R$.

Now, $B_{H_2}(0; R)$ is compact by the finite-dimensionality of H_2, whence the continuous function g attains its minimum on $B_{H_2}(0; R)$ at, say, y_0. What we have shown above shows that g attains its global minimum at y_0. □

We must show that $g(y_0) \leq 1$. To this end, we let

$$J = \left\{ n \in \{1, \ldots, N\} : \frac{\|y_0 - y_n\|}{\|x_0 - x_n\|} = g(y_0) \right\}.$$

We let G be an H_1-valued random variable with distribution $\mathbb{P}[G = x_n] = \alpha_n$ for all $n \in J$, where $\{\alpha_n\}_{n\in J}$ is the collection of scalars given in the claim below.

Claim. *We can write y_0 as a convex combination of $\{y_n\}_{n\in J}$, viz., there exist $\{\alpha_n\}_{n\in J} \subseteq [0,\infty)$ such that $\sum_{n\in J}\alpha_n = 1$ and $\sum_{n\in J}\alpha_n y_n = y_0$.*

Proof of claim. By the definition of g,

$$\{1,\dots,N\} \setminus J = \left\{ n : \frac{\|y_0 - y_n\|}{\|x_0 - x_n\|} < g(y_0) \right\},$$

and this shows that y_0 belongs to the convex hull of $\{y_n\}_{n\in J}$. Indeed, if $y_0 \notin \mathrm{Conv}(\{y_n\}_{n\in J})$, then the hyperplane separation theorem furnishes a hyperplane H that separate y_0 and $\mathrm{Conv}(\{y_n\}_{n\in J})$. Moving y_0 closer to H, we obtain a point y_* such that $\|y_* - y_n\| < \|y_0 - y_n\|$ for all $1 \le nN$. This, in particular, implies that

$$\frac{\|y_* - y_n\|}{\|x_0 - x_n\|} < g(y_0)$$

for all $1 \le n \le N$, contradicting the minimality of g at y_0. □

We define $h : \{x_1,\dots,x_N\} \to \{y_1,\dots,y_N\}$ by setting $h(x_n) = y_n$. We shall show that

$$g(y_0)^2\mathbb{E}[\|G - x_0\|^2] \le \mathbb{E}[\|G - \mathbb{E}G\|^2] \le \mathbb{E}[\|G - x_0\|^2], \tag{3.6}$$

which implies that $g(y_0) \le 1$. To this end, we first observe that

$$\begin{aligned}
\mathbb{E}\left[\|G - x_0\|^2 - \|G - \mathbb{E}G\|^2\right] =&\mathbb{E}\left[\|G\|^2 - 2\langle G, x_0\rangle + \|x_0\|^2\right] \\
&- \mathbb{E}\left[\|G\|^2 - 2\langle G, \mathbb{E}G\rangle + \|\mathbb{E}G\|^2\right] \\
=&\|x_0\|^2 - 2\mathbb{E}[\langle G, x_0\rangle] + 2\mathbb{E}[\langle G, \mathbb{E}G\rangle] - \|\mathbb{E}G\|^2 \\
=&\|x_0\|^2 - 2\langle \mathbb{E}G, x_0\rangle + 2\langle \mathbb{E}G, \mathbb{E}G\rangle - \|\mathbb{E}G\|^2 \\
=&\|x_0\|^2 - 2\langle \mathbb{E}G, x_0\rangle + \|\mathbb{E}G\|^2 \\
=&\|x_0 - \mathbb{E}G\|^2 \\
\ge&0,
\end{aligned}$$

whence the right half of (3.6) holds.

To establish the left half of (3.6), we observe that

$$g(y_0)\|G - x_0\| = \|h(G) - y_0\|$$

by the definitions of G and J. Therefore,

$$g(y_0)^2\mathbb{E}[\|G - x_0\|^2] = \mathbb{E}[\|h(G) - y_0\|^2].$$

Since

$$\mathbb{E}[h(G)] = \sum_{n\in J}\mathbb{E}[h(1_{\{G=x_n\}})] = \sum_{n\in J}\mathbb{P}[G = x_n]y_n = \sum_{n\in J}\alpha_n y_n = y_0,$$

we see that

$$\mathbb{E}[\|h(G) - y_0\|^2] = \mathbb{E}[\|h(G) - \mathbb{E}[h(G)]\|^2].$$

Combining what we have shown so far with the claim below, we obtain the identity

$$g(y_0)^2\mathbb{E}[\|G - x_0\|^2] = \mathbb{E}[\|h(G)\|^2 - \|\mathbb{E}h(G)\|^2]. \tag{3.7}$$

Claim. *Let H be a finite-dimensional Hilbert space and X an H-valued random variable. Then, for each, $x \in H$,*

$$\mathbb{E}[\|X - \mathbb{E}X\|^2] = \mathbb{E}[\|X\|^2 - \|\mathbb{E}X\|^2]. \tag{3.8}$$

Proof of claim. Observe that

$$\begin{aligned}\mathbb{E}\left[\|X-\mathbb{E}X\|^2\right] &= \mathbb{E}\left[\|X\|^2+\|\mathbb{E}X\|^2-2\langle X,\mathbb{E}X\rangle\right]\\ &= \mathbb{E}\left[\|X\|^2\right]+\mathbb{E}\,\|\mathbb{E}X\|^2-2\langle \mathbb{E}X,\mathbb{E}X\rangle\\ &= \mathbb{E}\left[\|X\|^2\right]+\mathbb{E}\left[\|\mathbb{E}X\|^2\right]-2\|\mathbb{E}X\|^2\\ &= \mathbb{E}\left[\|X\|^2\right]+\mathbb{E}\,\|\mathbb{E}X\|^2-2\mathbb{E}\left[\|\mathbb{E}X\|^2\right]\\ &= \mathbb{E}\left[\|X\|^2-\|\mathbb{E}X\|^2\right],\end{aligned}$$

as was to be shown. □

We now take an indepednent, identically distributed copy G' of G. This, in particular, implies that $h(G)$ and $h(G')$ are independent, identically distributed random variables. The claim below, along with (3.4), implies that

$$\mathbb{E}[\|h(G)\|^2-\|\mathbb{E}h(G)\|^2]=\frac{1}{2}\mathbb{E}\left[\|h(G)-h(G')\|^2\right]\leq\frac{1}{2}\mathbb{E}\left[\|G-G'\|^2\right]. \tag{3.9}$$

Claim. *If X and X' are independent, identically distributed random variables, then*

$$2\mathbb{E}\left[\|X\|^2-\|\mathbb{E}X\|^2\right]=\mathbb{E}\left[\|X-X'\|^2\right]. \tag{3.10}$$

Proof of claim. By independence, $\langle\mathbb{E}\langle X,X'\rangle=\langle\mathbb{E}X,\mathbb{E}X'\rangle$, and so

$$\begin{aligned}2\mathbb{E}\left[\|X\|^2-\|\mathbb{E}X\|^2\right] &= 2\mathbb{E}\left[\|X\|^2\right]-\|\mathbb{E}X\|^2\\ &= \left(\mathbb{E}\left[\|X\|^2\right]+\mathbb{E}\left[\|X'\|^2\right]-2\langle\mathbb{E}X,\mathbb{E}X'\rangle\right)\\ &= \mathbb{E}\left[\|X\|^2+\|X'\|^2-2\langle X,X'\rangle\right]\\ &= \mathbb{E}\left[\|X-X'\|^2\right],\end{aligned}$$

as was to be shown. □

(3.7) and (3.9) imply that

$$g(y_0)^2\mathbb{E}\left[\|G-x_0\|^2\right]\leq\frac{1}{2}\mathbb{E}\left[\|G-G'\|^2\right].$$

Applying Claim (3.10) to the right-hand side, we obtain the inequality

$$g(y_0)^2\mathbb{E}\left[\|G-x_0\|^2\right]\leq\mathbb{E}\left[\|G\|^2-\|\mathbb{E}G\|^2\right].$$

We now invoke Claim (3.8) to conclude that

$$g(y_0)^2\mathbb{E}\left[\|G-x_0\|^2\right]\leq\mathbb{E}\left[\|G-\mathbb{E}G\|^2\right],$$

which is the left half of (3.6). This completes the proof of the lemma under the finiteness assumptions. □

Step 2. The lemma holds whenever $|I|<\infty$.

We now assume that H_1 and H_2 are Hilbert spaces of arbitrary dimensions, while continuing to suppose that $I=\{1,\ldots,N\}$. Let $\tilde{H}_1=\operatorname{span}\{x_1,\ldots,x_N\}$ and $\tilde{H}_2=\operatorname{span}\{y_1,\ldots,y_N\}$. which are finite-dimensional. Define $f:\{x_1,\ldots,x_N\}\to\tilde{H}_2$ by setting $f(x_n)=y_n$ for all $1\leq n\leq N$. Hypothesis (3.4) implies that f is 1-Lipschitz.

Step 1 implies that the Kirszbraun extension theorem holds for Lipschitz maps of finite image on finite-dimensional Hilbert spaces, and so we can find a 1-Lipschitz extension $F:\tilde{H}_1\to\tilde{H}_2$ of f. The standard inclusion map $\iota:\tilde{H}_2\to H_2$ is an isometry, and so $\iota\circ F:\tilde{H}_1\to H_2$ is 1-Lipschitz as well. Since the orthogonal projection $\rho : H_1 \to \tilde{H}_1$ is 1-Lipschitz as well, we see that $\tilde{F} = \iota \circ F \circ \rho$ is a 1-Lipschitz extension of $\iota \circ F$. See the diagram below:

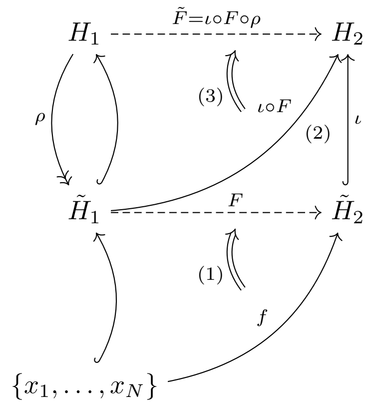


FIGURE 4. (1) Extend; (2) Compose; (3) Extend

Fix $x_* \in \bigcap_{1\le n\le N} B_{H_1}(x_n, r_n)$. By the Lipschitz condition on $\tilde{F}$, we have the inequality

$$\|\tilde{F}(x_*) - y_n\| = \|\tilde{F}(x_*) - \tilde{F}(x_n)\| \le \|x_* - x_n\| \le r_n$$

holds for all $1 \le n \le N$. It follows that $\tilde{F}(x_*) \in \bigcap_{1\le N} B_{H_2}(y_n, r_n)$, as was to be shown. □

***Step 3. The lemma holds in all cases*[2].**

Let us now lift all finiteness assumptions and assume that H_1, H_2, and I are arbitrary. Define a function $f : \{x_\beta\}_{\beta\in I} \to H_2$ by setting $f(x_\beta) = y_\beta$, which is 1-Lipschitz by Hypothesis (3.4).

We fix $\beta_0 \in I$ and set, for each $x \in H_1$,

$$B_x = \{y \in H_2 : \|y\| \le \|y_{\beta_0}\| + \|x - x_{\beta_0}\|\},$$

which is a ball in H_2. The Banach-Alaoglu theorem implies that B_x is weakly compact. The topological product $B = \prod_{x\in H_1} B_x$ is then compact with respect to the weak topology on H_2; this is a consequence of Tychonoff's theorem. Note that each element of B is a function $g : H_1 \to H_2$ that maps $x \in H_2$ into B_x.

For each finite subset A of H_1, we define

$$\mathcal{G}_A = \left\{g \in B : g \text{ is a 1-Lipschitz extension of } f|_{\{x_\beta\}_{\beta\in I}\cap A} \text{ on } A\right\}.$$

Note that the intersection

$$\mathcal{G} = \bigcap_{\substack{A\subseteq H_1\\ |A|<\infty}} \mathcal{G}_A$$

is a set of 1-Lipschitz extensions of f on H_1. We shall show that each $\mathcal{G}_A$ is a nonempty closed set and that

$$\mathscr{G} = \{\mathcal{G}_A : A \text{ is a finite subset of } H_1\}$$

[2] Step 3 is taken from Fremlin's notes [Fre].

satisfies the finite intersection property. Since B is compact, this would imply that the intersection $\mathscr{G}$ is nonempty, as was to be shown.

Step 2 implies that the Kirszbraun extension theorem holds for Lipschitz maps of finite image. If A is finite, then the set $\{x_\beta\}_{\beta\in I}\cap A$ is also finite, and so $\operatorname{im} f|_{\{x_\beta\}_{\beta\in I}}$ is finite. Therefore, there is a 1-Lipschitz extension of $f|_{\{x_\beta\}_{\beta\in I}\cap A}$, whence it follows that $\mathcal{G}_A$ is nonempty.

This, in particular, shows that $\mathscr{G}$ satisfies the finite intersection property. Indeed, if $A_1,\ldots,A_N$ are finite subsets of H_1, then $A = A_1\cup\cdots\cup A_N$ is finite. Moreover, $\mathcal{G}_A\subseteq\mathcal{G}_{A_n}$ for all $1\le n\le N$, hence

$$\bigcap_{n=1}^{N}\mathcal{G}_{A_n}\supseteq\mathcal{G}_A,$$

which is nonempty by what we have shown above.

It remains to show that $\mathcal{G}_A$ is closed for each finite subset A of H_1. To this end, we define two subsets of B as follows:

$$\begin{aligned}\mathcal{G}'_A &= \{g\in B : g \text{ is a (possibly discontinuous) extension of } f|_{\{x_\beta\}_{\beta\in I}\cap A}\}\\ \mathcal{G}''_A &= \{g\in B : g \text{ is 1-Lipschitz on } A\}.\end{aligned}$$

Since

$$\mathcal{G}_A=\mathcal{G}'_A\cap\mathcal{G}''_A,$$

it suffices to show that both $\mathcal{G}'_A$ and $\mathcal{G}''_A$ are closed in B.

Let us recall that the evaluation function $\pi_x : B\to B_x$, given by $\pi_x(g)=g(x)$, is continuous for each $x\in H_1$. Now, $\{x_\beta\}_{\beta\in I}\cap A$ is a finite set, and so $A' = \{f(x)\}_{x\in\{x_\beta\}_{\beta\in I}\cap A}$ is a closed subset of H_2. Therefore,

$$\pi_x^{-1}(A')=\{g\in B : g(x)=f(x)\}$$

is a closed subset of B, provided that $x\in\{x_\beta\}_{\beta\in I}\cap A$. It follows that

$$\mathcal{G}'_A=\bigcap_{x\in\{x_\beta\}_{\beta\in I}\cap A}\pi_x^{-1}(A')$$

is a closed subset of B.

We now show that $\mathcal{G}''_A$ is a closed subset of B. To this end, we recall that the evaluation function π_x, now considered as a map into H_2, is continuous with respect to the weak topology on H_2. Furthermore, the inner-product evaluation map $\varphi_a : H_2\to\mathbb{R}$ given by $\varphi_a(b)=\langle b,a\rangle_{H_2}$ is continuous with respect to the strong topology on H_2, regardless of the choice of $a\in H_2$. Fixing $x,y\in H_1$ and $a\in H_2$, we see that

$$\begin{aligned}[\varphi_a\circ(\pi_x-\pi_y)]^{-1}([\|x-y\|,\infty)) &= (\pi_x-\pi_y)^{-1}(\{b\in H_2 : \langle b,a\rangle_{H_2}\le\|x-y\|\})\\ &= \{g\in B : \langle g(x)-g(y),a\rangle_{H_2}\le\|x-y\|\}\end{aligned}$$

is a closed subset of B. By the Riesz representation theorem, there exists a point $\|a_{x,y}\|\le 1$ in H_2 such that $\langle g(x)-g(y),a\rangle_{H_2}=\|g(x)-g(y)\|_{H_2}$. It now follows that

$$\mathscr{G}''_A=\bigcap_{\substack{a\in H_2\\ \|a\|\le 1}}\{g\in B : \langle g(x)-g(y),a\rangle_{H_2}\le\|x-y\|\},$$

whence $\mathscr{G}''_A$ is a closed subset of B.

This concludes the proof of Lemma 3.3, thereby establishing the Kirszbraun extension theorem as well. □

Remark. Lemma 3.3 serves as a motivation for a geometric characterization of extendibility, referred to as the *binary intersection property*. See Chapter 1 of [BL00] for details.

Remark. If $H_1 = H_2$, then the contraction hypothesis in Lemma 3.3 amounts to rearranging the balls $\{B(x_\beta, r_\beta)\}$ in a way that the centers of the balls are closer to each other. If, in addition, $\dim H_1 = \dim H_2 < \infty$, then it is natural to inquire how this act of rearrangement changes the volume of the intersection. In fact, we might guess that

$$\mathrm{Vol}\left(\bigcap_\beta B(y_\beta, r_\beta)\right) \geq \mathrm{Vol}\left(\bigcap_\beta B(x_\beta, r_\beta)\right).$$

This is a conjecture of Kneser and Poulsen, which, surprisingly enough, is still open in dimensions 3 and higher as of March 2014. See §4 for details.

With the Kirszbraun extension theorem at our disposal, we can now make progress on the problem of uniform extendibility of Hilbert-valued functions, considered in Examples 2.1 and 2.2.

Corollary 3.11. *Let (X, d) be a metric space, A a subset of X, and H a Hilbert space. If, for some $0 < \alpha \leq 1/2$, the function $f : A \to H$ is α-Hölder, then there exists an α-Hölder extension $F : X \to H$ of f such that $\|F\|_{\mathscr{C}^{0,\alpha}} = \|f\|_{\mathscr{C}^{0,\alpha}}$.*

The above result is sharp, as demonstrated in Example 2.8.

Proof. We observe first that it suffices to prove the Corollary for the $\alpha = 1/2$. Indeed, if $\alpha < 1/2$, then f can be considered as an $(1/2)$-Hölder function from $(A, d^{2\alpha})$ into H. Taking a $(1/2)$-Hölder extension $F : (X, d^{2\alpha}) \to H$ of f, we see that $F : (X, d) \to H$ is an α-Hölder extension of $f : (A, d) \to H$. We remark that the restriction $\alpha < 1/2$ is necessary for this method, as d^β is not a metric if $\beta > 1$.

Let us now suppose that $\alpha = 1/2$. By Zorn's lemma, it suffices to extend f by one point. We therefore assume that A is a proper subset of X, and fix $x_* \in X \smallsetminus A$. Moreover, by considering $(\|f\|_{\mathscr{C}^{0,1/2}})^{-1} f$ if necessary, we assume without loss of generality that $\|f\|_{\mathscr{C}^{0,1/2}} = 1$.

Consider the families of balls

$$\mathscr{B}_1 = \left\{ B_{l_2(X)}\left(\sqrt{d(x_*, x)}\, x; \sqrt{d(x_*, x)}\right) \right\}_{x \in A};$$
$$\mathscr{B}_2 = \left\{ B_H\left(f(x); \sqrt{d(x_*, x)}\right) \right\}_{x \in A}.$$

The $(1/2)$-Hölder continuity of f and the triangle inequality yield the following estimate:

$$\begin{aligned}\|f(x) - f(y)\|_H &\leq \sqrt{d(x, y)} \\ &\leq \sqrt{d(x, x_*) + d(x_*, y)} \\ &= \left\| \sqrt{d(x_*, x)}\, x - \sqrt{d(x_*, y)}\, y \right\|_{l_2(X)}.\end{aligned}$$

By construction, $x_* \in \bigcap_{B \in \mathscr{B}_1} B$. We now invoke Lemma 3.3 to establish the existence of a point $y_* \in \bigcap_{B \in \mathscr{B}_2} B$. It now suffices to note that the function $\tilde{f} : A \cup \{x_*\} \to H$ defined by setting

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in A \text{ and} \\ y_* & \text{if } x = x_* \end{cases}$$

is a (1/2)-Hölder extension of f. □

We conclude with a result on *uniform* extension of Hilbert-valued functions.

Theorem 3.12 (Grünbaum–Zarantonello extension theorem, 1968; [GZ68]). *Let H_1 and H_2 be a Hilbert space, A a subset of H_1, and $f : A \to H_2$ a uniformly continuous map. There exists a uniformly continuous extension $F : H_1 \to H_2$ of f if and only if f admits a nondecreasing subadditive function $\omega : [0, \infty) \to [0, \infty)$ such that $\lim_{t \to 0^+} \omega(t) = 0$ and $\omega_f \leq \omega$*

4. Further Results: The Kneser–Poulsen conjectures

Fix a finite-dimensional Euclidean space $\mathbb{R}^d$. A finite set of points $\{y_1, \ldots, y_N\} \subseteq \mathbb{R}^d$ is a *contraction* of another set of points $\{x_1, \ldots, x_N\} \subseteq \mathbb{R}^d$ if

$$\|y_m - y_n\| \leq \|x_m - x_n\|$$

for all $1 \leq m, n \leq N$. If $\{y_1, \ldots, y_N\}$ is a contraction of $\{x_1, \ldots, x_N\}$, then, for each collection $\{r_1, \ldots, r_N\}$ of nonnegative numbers, the Kirszbraun extension theorem (Chapter 1, Lemma 3.3) shows that

$$\bigcap_{n=1}^{N} B(y_n, r_n) \neq \varnothing$$

whenever $\bigcap_{n=1}^{N} B(x_n, r_n) \neq \varnothing$. Intuitively, however, more should be true. Indeed, it is natural to speculate that the volume of the intersection should increase as we push the spheres together. Formally, we are led to the following conjecture:

Conjecture 4.1 (Kneser–Poulsen conjecture 1). *Let $\{r_1, \ldots, r_N\}$ be a collection of nonnegative numbers and $\{x_1, \ldots, x_N\}$ and $\{y_1, \ldots, y_N\}$ two subsets of $\mathbb{R}^d$. If $\{y_1, \ldots, y_N\}$ is a contraction of $\{x_1, \ldots, x_N\}$, then*

$$\text{Vol}\left(\bigcap_{n=1}^{N} B(y_n, r_n)\right) \geq \text{Vol}\left(\bigcap_{n=1}^{N} B(x_n, r_n)\right).$$

The above was conjectured independently by E. T. Poulsen and M. Kneser. In [Pou54], Poulsen posed the following question, closely related to Conjecture 4.1:

Conjecture 4.2 (Poulsen). *Let $r > 0$, and let $\{x_1, \ldots, x_N\}$ and $\{y_1, \ldots, y_N\}$ be two subsets of $\mathbb{R}^d$. If $\{y_1, \ldots, y_N\}$ is a contraction of $\{x_1, \ldots, x_N\}$, then*

$$\text{Vol}\left(\bigcup_{n=1}^{N} B(y_n, r)\right) \leq \text{Vol}\left(\bigcup_{n=1}^{N} B(x_n, r)\right).$$

Kneser, in [Kne55], makes the same conjecture after establishing the following theorem:

Theorem 4.3 (Kneser, 1955; [Kne55]). *Let $r > 0$, and let $\{x_1, \ldots, x_N\}$ and $\{y_1, \ldots, y_N\}$ be two subsets of $\mathbb{R}^d$. If $\{y_1, \ldots, y_N\}$ is a contraction of $\{x_1, \ldots, x_N\}$, then*

$$\operatorname{Vol}\left(\bigcup_{n=1}^{N} B(y_n, r)\right) \leq 3^d \operatorname{Vol}\left(\bigcup_{n=1}^{N} B(x_n, r)\right).$$

The theorem is proved by investigating the subcollection $\{B(y_n, r)\}_{n\in I}$, where I is the maximal subset of $\{1, \ldots, N\}$ such that $\|x_n - x_m\| \geq 2r$ for all distinct $n, m \in I$.

Both Conjecture 4.2 and Theorem 4.3 are special cases of the following conjecture, which is similar but non-equivalent to Conjecture 4.1:

Conjecture 4.4 (Kneser–Poulsen conjecture 2). *Let $\{r_1, \ldots, r_N\}$ be a collection of nonnegative numbers and $\{x_1, \ldots, x_N\}$ and $\{y_1, \ldots, y_N\}$ two subsets of $\mathbb{R}^d$. If $\{y_1, \ldots, y_N\}$ is a contraction of $\{x_1, \ldots, x_N\}$, then*

$$\operatorname{Vol}\left(\bigcup_{n=1}^{N} B(y_n, r_n)\right) \leq \operatorname{Vol}\left(\bigcup_{n=1}^{N} B(x_n, r_n)\right).$$

As of March 2014, the Kneser–Poulsen conjectures are resolved only in dimension 2. Integral to the solution of the two-dimensional conjectures and the various partial solutions to the higher-dimensional conjectures is the following notion:

Definition 4.5. Let $\{x_1, \ldots, x_N\}$ and $\{y_1, \ldots, y_N\}$ be two subsets of $\mathbb{R}^d$. The set $\{y_1, \ldots, y_N\}$ is a *continuous (respectively, analytic) contraction* of $\{x_1, \ldots, x_N\}$ if there exist continuous (respectively, analytic) curves $\gamma_1, \ldots, \gamma_N : [0, 1] \to \mathbb{R}^d$ such that

- $\gamma_n(0) = x_n$ for all $1 \leq n \leq N$,
- $\gamma_n(1) = y_n$ for all $1 \leq n \leq N$, and
- $\|\gamma_n(t) - \gamma_m(t)\|$ is monotonically decreasing on $[0, 1]$ for all $1 \leq m, n \leq N$.

Before the 2002 solution to the two-dimensional Kneser–Poulsen conjectures was published, there were two main lines of attacking the problems: investigating continuous contractions in dimension 2, and studying arbitrary contractions of a small number of points in high dimensions. In [BC02], K. Bezdek and R. Connelly combine both ideas to settle the two-dimensional conjectures. Roughly, it is shown that d-dimensional contractions can be represented by $2d$-dimensional analytic contractions, and that the volumes of $2d$-dimensional spheres can be related to the volumes of $(2d-2)$-dimensional slices. See [BC02] or Chapters 5 and 11 of [Bez10] for the full proof. The introduction to [BC02] also contains references to several partial results.

5. Appendix: Graph Theory

In this section, we review graph-theoretic notions that are used throughout the notes. The primary reference for this section is [Die10].

Definition 5.1. A *graph* G is an ordered pair (V, E) of a nonempty set V and a collection E of two-element subsets of V. An element of V is called a *vertex* or a *node*, and the cardinality $|V|$ of V is the *order* of G. An element $\{x, y\}$ of E, is called an *edge between x and y*, and we write xy to denote $\{x, y\}$; if $xy \in E$, then we say that $x, y \in V$ are *adjacent*. Two graphs (V_1, E_1) and (V_2, E_2) are *isomorphic* if there exists a bijection $\varphi : V_1 \to V_2$ such that $xy \in E_1$ if and only if $\varphi(x)\varphi(y) \in E_2$.

Example 5.2 (Complete graphs). A graph is *complete* if every distinct pair of vertices is adjacent to each other. We see at once that two complete graphs with the same order are isomorphic. We write K_n to denote the complete graph of order n.

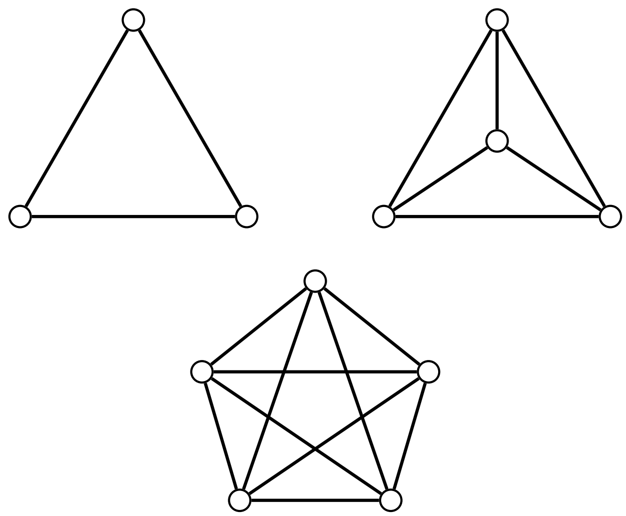


FIGURE 5. K^3, K^4, and K^5

Definition 5.3. Given a graph $G = (V, E)$ and two vertices $x, y \in V$, a *path from x to y* is a finite subgraph P of G of the form

$$V(P) = \{x_0, \dots, x_n\} \quad \text{and} \quad E(P) = \{x_0x_1, \dots, x_{n-1}x_n\}, \tag{5.4}$$

where $x_0 = x$ and $x_n = y$.

Definition 5.5. A *metrized graph* is a graph (V, E) with a weight function $w : E \to (0, \infty)$ that assigns the length of each edge, and a distance function $d : V \times V \to [0, \infty)$ that determines the distance between two vertices, given by

$$d(x, y) = \inf \sum_{k=1}^{n} w(x_{k-1}x_k),$$

where the infimum is taken over all paths (5.4) from x to y.

Proposition 5.6 (Graph metric)**.** *Let (V, E, d) be a metrized graph. If $d(x, y)$ is finite for all $\{xy\} \in E$, then d defines a metric on the set V, called a* graph metric, *or a* shortest path metric with weight w, *on the graph (V, E).* □

Example 5.7. Take K^5 and label the nodes counterclockwise. Here are some possible paths from node 1 to node 3.

If we introduce a weight function that assigns the length of 1 to each edge, then we see that the length from node 1 to node 3 is 1. In fact, the length between two distinct nodes of K^5 is always 1, whence our graph metric turns K^5 into the discrete metric space on five elements.

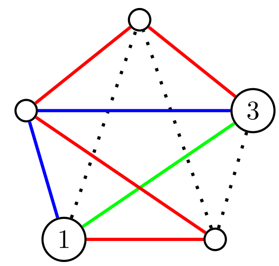


FIGURE 6. Paths from 1 to 3

6. Appendix: Fourier Analysis on Boolean Hypercubes

One of the most frequently used tools in mathematics is the *Fourier transform*, which decomposes a given function into a superposition of more basic functions, often symmetric in nature. Classical Fourier analysis provides us with two versions of the Fourier transform: the *Fourier coefficient operator* that produces the Fourier coefficients

$$\hat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx$$

of a 2π-periodic function $f : [-\pi, \pi] \to \mathbb{C}$, and the *Fourier transform operator* that produces the Fourier transform

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i x\xi}\,dx$$

of a non-periodic function $f : \mathbb{R} \to \mathbb{R}$. These are invertible, in the sense that we have the *Fourier series*

$$f(x) = \sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx}$$

and the *inverse Fourier transform*

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi)e^{2\pi i\xi x}\,d\xi.$$

Furthermore, both versions satisfy useful L^2-identities: *Parseval's theorem*

$$\sum_{n\in\mathbb{Z}} |\hat{f}(n)|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2\,dx$$

and *Plancherel's theorem*

$$\int_{-\infty}^{\infty} |f(x)|^2\,dx = \int_{-\infty}^{\infty} |\hat{f}(\xi)|^2\,d\xi.$$

In what sense are these two versions instantiations of one unified concept, *the* Fourier transform? To make the connections clearer, we transport 2π-periodic functions on $\mathbb{R}$ onto the circle group $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ via the change-of-variables map $x \mapsto e^{ix}$. With the usual multiplication operation as group operation, $\mathbb{T}$ is an abelian group. Furthermore, the restriction of the Euclidean topology on $\mathbb{T}$ turns it into a compact topological group[3]. Carrying over the one-dimensional Lebesgue measure onto $\mathbb{T}$, we end up with a *Haar measure* σ on $\mathbb{T}$, viz., $\sigma(zE) =$

[3] We recall that a *topological group* is a group G with a topology that renders the group operation map $G \times G \to G$ and the inverse map $G \to G$ continuous.

$\sigma(E)$ for each $z \in \mathbb{T}$ and every μ-measurable subset of $\mathbb{T}$. Compare this with $\mathbb{R}$, a locally compact abelian group on which the Lebesgue measure is a Haar measure.

Given a locally compact abelian group G with a Haar measure μ, we develop *Fourier analysis on* G as follows. Firstly, we consolidate the *continuous characters* of G, which are continuous functions $\chi : G \to \mathbb{C}$ such that $\chi(gh) = \chi(g)\chi(h)$ for all $g, h \in G$. The collection of all continuous characters of G is called the *dual group* of G, denoted by $\widehat{G}$: note, for example, that $\widehat{\mathbb{T}} = \{e^{inx} : n \in \mathbb{Z}\}$ and $\widehat{\mathbb{R}} = \{e^{2\pi i x \xi} : \xi \in \mathbb{R}\}$. The *Fourier transform* is then defined to be the integral transform $\mathscr{F} : L^1(G) \to \mathscr{C}_0(\hat{G})$ that sends $f \in L^1(G)$ to

$$(\mathscr{F}f)(\xi) = \hat{f}(\xi) = \int_G f(x)\overline{\xi(x)}\, dx.$$

With this definition, we can endow $\widehat{G}$ with an appropriate Haar measure, called the *dual measure*, that renders the *Fourier inversion formula*

$$f(x) = \int_{\widehat{G}} \hat{f}(\xi)\xi(x)\, d\xi$$

and *Plancherel's theorem*

$$\int_G |f(x)|^2\, dx = \int_{\widehat{G}} |\hat{f}(\xi)|^2\, d\xi$$

true. This is the *abstract theory of the Fourier transform* that unites the two versions of the Fourier transform studied in classical Fourier analysis.

The goal of this section is to study another concrete instantiation of the abstract theory. Specifically, we develop Fourier analysis on the *Boolean hypercube*, which is a space of *bit sequences*, i.e., sequences whose terms are selected from a two-element set. As the name suggests, the Boolean hypercube is ubiquitous in theoretical computer science. It also serves as an abstract model for Bernoulli trials in probability theory, and as an oft-used counterexample in metric geometry.

Formally, we take the *N-dimensional Boolean hypercube* to be the N-fold direct product

$$\mathbb{Z}_2^N = \{x = (x_1, \ldots, x_N) : x_i \in \mathbb{Z}_2\}$$

of the cyclic group of order 2 $\mathbb{Z}_2$ with the normalized counting measure $\mathbb{P}_N = 2^{-N}\#$. For the purposes of developing Fourier analysis on Boolean hypercubes, it is be useful to consider the infinite Boolean hypercube

$$\mathbb{Z}_2^\infty = \{x = (x_n)_{n=1}^\infty : x_i \in \mathbb{Z}_2\},$$

which is the $\mathbb{N}$-fold direct product of $\mathbb{Z}_2$ with the probability measure given by the following construction:

Lemma 6.1. *Let $\{(\Omega_j, \Sigma_j, \mathbb{P}_j)\}_{j \in \mathbb{N}}$ be a collection of probability spaces, and define a* cylinder set *E as a subset of $\prod_j X_j$ of the form*

$$\{x = (x_j) : x_j \in E_j \in \mathcal{M}_j \text{ and } E_j = X_j \text{ for almost all } j\}.$$

There exists a product measure $\mathbb{P}$ on $\Omega = \prod_j \Omega_j$ such that $\mathbb{P}(E) = \prod_j \mathbb{P}_j(E_j)$ for all cylinder sets E.

Proof of lemma. It is enough to show that $\mathbb{P}$ is countably additive for cylinder sets that are disjoint unions of cylinder sets. The Carathéodory extension theorem takes care of the rest.

Let $E = \prod_{j=1}^{\infty} E^j$ be a cylinder set and suppose that E can be written as the disjoint union of a countable collection $\{E_n = \prod_{j=1}^{\infty} E_n^j\}_{n\in\mathbb{N}}$ of cylinder sets. We assume without loss of generality that $E \neq \prod_{j=1}^{\infty} \Omega_j$ and find $k \in \mathbb{N}$ such that $E^j = \Omega_j$ for all $j > k$ and that $E^k \neq \Omega_j$. Taking unions if necessary, we conclude from this assumption that $E_n^j = \Omega_j$ for each $j > k$ and every $n \in \mathbb{N}$. Observe that

$$\mathbb{P}(E) = \mathbb{P}_1(E^1)\cdots\mathbb{P}_k(E^k)\cdot\prod_{j=k+1}^{\infty}\mathbb{P}_j(E^j) = \mathbb{P}_1(E^1)\cdots\mathbb{P}_k(E^k).$$

By the argument for the k-fold product case, we have

$$\mathbb{P}(E) = \sum_{n=1}^{\infty}\mathbb{P}_1(E_n^1)\cdots\mathbb{P}_k(E_n^k) = \mathbb{P}(E) = \sum_{n=1}^{\infty}\mathbb{P}_1(E_n^1)\cdots\mathbb{P}_k(E_n^k)\cdot\prod_{j=k+1}^{\infty}\mathbb{P}_j(E_n^j),$$

and we have countable additivity. □

We now consider $\mathbb{Z}_2$ to be the multiplicative group $\{-1, 1\}$ and define

$$\mathfrak{r}_n(x) = x_n$$

for each $x \in \mathbb{Z}_2^{\infty}$ and every $n \in \mathbb{N}$. We see at once that $(\mathfrak{r}_n)_{n=1}^{\infty}$ are identically distributed random variables on $\mathbb{Z}_2^{\infty}$. We now prove that $(\mathfrak{r}_n)_{n=1}^{\infty}$ are identically distributed, thereby showing that $(\mathfrak{r}_n)_{n=1}^{\infty}$ is an example of a *Rademacher sequence.*

Definition 6.2. A *Rademacher sequence* is a sequence $(\varepsilon_n)_{n=1}^{\infty}$ of independent, identically distributed random variables on some probability space $(\Omega, \Sigma, \mathbb{P})$ with the distribution $\mathbb{P}(\varepsilon_n = 1) = \mathbb{P}(\varepsilon_n = -1) = \frac{1}{2}$.

Independence of $(\mathfrak{r}_n)_{n=1}^{\infty}$ is a consequence of the following lemma:

Lemma 6.3. *Let $(X, \mathbb{P})$ be the product of a collection $\{(X_n, \mathbb{P}_n)\}_{n\in\mathbb{N}}$ of probability spaces. If a sequence of random variables on $(f_n)_{n=1}^{\infty}$ admits random variables $F_n : X_n \to \mathbb{R}$ such that $f_n(x) = F_n(x_n)$ for all $n \in \mathbb{N}$ and $x \in X$, then $(f_n)_{n=1}^{\infty}$ is mutually independent.*

Proof of lemma. Let $(B_n)_{n=1}^{\infty}$ be a sequence of Borel subsets of $\mathbb{R}$ and let $E_n = \{x : f_n(x) \in B_n\}$ and $E_n' = \{x_n : F_n(x_n) \in B_n\}$. Then $E_n = \{x : x_n \in E_n'\}$, so that E_n is a cylinder set with $\mathbb{P}(E_n) = \mathbb{P}_n(E_n')$. By the construction of the infinite product measure $\mathbb{P}$, we see that

$$\mathbb{P}\left(\bigcap_{n=1}^{N} E_n\right) = \prod_{n=1}^{N}\mathbb{P}_n(E_n') = \prod_{n=1}^{N}\mathbb{P}(E_n),$$

whence sending $N \to \infty$ yields the desired result. □

Observe that $(\mathfrak{r}_n)_{n=1}^{\infty}$ is not an orthonormal basis of $L_2(\mathbb{Z}_2^{\infty})$. Indeed,

$$\langle \mathfrak{r}_n, 1\rangle_{L_2(\mathbb{Z}_2^{\infty})} = \mathbb{E}[\mathfrak{r}_n] = 0$$

for all $n \in \mathbb{N}$. In particular, this implies that $(\mathfrak{r}_n)_{n=1}^{\infty}$ cannot serve as a Fourier basis of $L_2(\mathbb{Z}_2^{\infty})$. To remedy this issue, we collect all finite products of our Rademacher sequence, labeling them as follows. We set $\mathfrak{w}_0(t) = 1$ and set, for each $n = 2^{k_1} + \cdots + 2^{k_l}$ with $k_1 > \cdots > k_l$,

$$\mathfrak{w}_n(t) = \prod_{i=1}^{l}\mathfrak{r}_{k_i}(t).$$

The resulting sequence $(w_n)_{n=0}^\infty$ is the sequence of *Walsh functions*.

We remark that the term *Walsh functions* in literature typically refers to an orthonormal basis $(w_n)_{n=0}^\infty$ in $L_2([0,1])$ closely related to $(\mathfrak{w}_n)_{n=0}^\infty$. To construct $(w_n)_{n=0}^\infty$, we recall that each real number in $[0,1]$ admits a unique binary expansion, so long as we adopt a convention to accept either the repeated 1's at the end or the finitary counterparts thereof via repeated 0's at the end, but not both. Then there is an injective mapping $D : [0,1] \to \mathbb{Z}_2^\infty$, where $N = \mathbb{Z}_2^\infty \smallsetminus [0,1]$ is a countable set. We now set

$$r_n(t) = \mathfrak{r}_n(D(t)),$$

so that r_n is a random variable on $[0,1]$, modulo a countable set. Observe that

$$r_1(t) = \begin{cases} 1 & \text{if } 0 < t < 1/2; \\ -1 & \text{if } 1/2 < t < 1. \end{cases}$$

Furthermore, if we take a 1-periodic extension of r_1 onto $\mathbb{R}$, then $r_n(t) = r_1(2^{n-1}t)$ on $[0,1]$. The functions r_n are referred to as the *Rademacher functions*, and the products

$$w_n(t) = \prod_{i=1}^{l} r_{k_i}(t).$$

are referred to as the *Walsh functions*.

It can be shown that $\mathscr{W}$ is an orthonormal basis of $L^2([0,1])$. In fact, the collection $\{w_n : n \in \mathbb{N}\}$ is precisely the dual group of $\mathbb{Z}_2^\infty$, and, with that, Fourier analysis on $\mathbb{Z}_2^\infty$ can be developed. As we have not provided rigorous justification of the abstract machinery sketched at the beginning of this section[4], we content ourselves with Fourier analysis on finite-dimensional Boolean hypercubes. This can be developed "by hand" using calculus and linear algebra.

We continue to consider $\mathbb{Z}_2^N$ as the N-fold product of $\{-1,1\}$. Following [O'D], we write $[N]$ to denote the set $\{1,\ldots,N\}$. For each $S \subseteq \{1,\ldots,n\}$, we have the the Walsh fucntion

$$\chi_S(x_1,\ldots,x_N) = \prod_{n \in S} x_n.$$

Independence of Rademacher functions implies that $\mathscr{W}_N = \{\chi_S : S \subseteq [N]\}$ is an orthonormal set in $L_2(\mathbb{Z}_2^N)$. Since $\dim = L_2(\mathbb{Z}_2^N) = 2^N = |\mathscr{W}_N|$, we see that $\mathscr{W}_N$ is an orthonormal basis of $\mathscr{W}_N$. The theory of finite-dimensional inner product spaces now guarantees that the *Fourier–Walsh expansion*

$$f(x) = \sum_{S \subseteq [N]} \hat{f}(S)\chi_S(x),$$

the uniqueness of the Fourier–Walsh expansion, and the *Parseval identity*

$$\langle f, f \rangle_{L_2(\mathbb{Z}_2^N)} = \sum_{S \subseteq [N]} \hat{f}(S)^2$$

are valid for all functions $f : \mathbb{Z}_2^N \to \mathbb{R}$. Likewise, simple computations show that foundational Fourier-analytic results such as the *convolution theorem*

$$f \hat{*} g(S) = \hat{f}(S)\hat{g}(S)$$

[4]See [Fol95] for a detailed treatment of the abstract theory, as well as many concrete instantiations thereof.

and the *Hausdorff–Young inequality*

$$\|\hat{f}\|_{L_{p'}(\mathbb{Z}_2^N)} \leq \|f\|_{L_p(\mathbb{Z}_2^N)},$$

where $1 \leq p \leq 2$ and $\frac{1}{p} + \frac{1}{p'} = 1$, continue to hold.

7. Appendix: Metric type

A defining property of a norm on a vector space is the *triangle inequality*

$$\left\| \sum_{n=1}^{N} x_n \right\| \leq \sum_{n=1}^{N} \|x_n\|.$$

Since the triangle inequality is insensitive to the choice of signs, we have the *randomized triangle inequality*

$$\mathbb{E} \left\| \sum_{n=1}^{N} \varepsilon_n x_n \right\| \leq \sum_{n=1}^{N} \|x_n\|,$$

where $(\varepsilon_n)_{n=1}^{\infty}$ is a Rademacher sequence (Definition 6.2). The l_p-norms on a finite sequence space are equivalent, and so we can replace the l_1-norm above by the l_p-norm to obtain the inequality

$$\left(\mathbb{E} \left\| \sum_{n=1}^{N} \varepsilon_n x_n \right\| \right)^{1/p} \leq C_{p,N} \left(\sum_{n=1}^{N} \|x_n\|^p \right)^{1/p},$$

where the constant $C_{p,N}$ depends only on the order p and the number N of vectors $x_1, \ldots, x_N$. As the dependence on N renders the inequality unwieldy as $N \to \infty$, we seek to remove the dependence. This leads us to the following definition:

Definition 7.1. A Banach space X is of *Rademacher type* p for some $1 \leq p \leq 2$ if there exists a constant C_p, depending only on p, such that the inequality

$$\left(\mathbb{E} \left\| \sum_{n=1}^{N} \varepsilon_n x_n \right\|^p \right)^{1/p} \leq C_p \left(\sum_{n=1}^{N} \|x_n\|^p \right)^{1/p} \tag{7.2}$$

holds for each finite sequence $(x_n)_{n=1}^{N}$, where $(\varepsilon_n)_{n=1}^{\infty}$ is a Rademacher sequence. The infimum of all such constants is denoted by $T_p(X)$.

An important property of Rademacher type is that it is an isomorphic invariant. Indeed, if X is of Rademacher type p, and if $T : Y \to X$ is an isomorphism of Banach spaces, then

$$\begin{aligned}
\left(\mathbb{E} \left\| \sum_{n=1}^{N} \varepsilon_n y_n \right\| \right)^{1/p} = \left(\mathbb{E} \left\| \sum_{n=1}^{N} \varepsilon_n T^{-1} x_n \right\| \right)^{1/p} &= \left(\mathbb{E} \left\| T^{-1} \sum_{n=1}^{N} \varepsilon_n x_n \right\| \right)^{1/p} \\
&\leq \left(\mathbb{E} \left[\|T^{-1}\| \left\| \sum_{n=1}^{N} \varepsilon_n x_n \right\| \right] \right)^{\frac{1}{p}} \\
&\leq \|T^{-1}\|^{\frac{1}{p}} T_p(X) \left(\sum_{n=1}^{N} \|x_n\|^p \right)^{\frac{1}{p}},
\end{aligned}$$

and so Y is of type p.

We note that this invariant is *local*: it only depends on finite-dimensional subspaces of the Banach space in question. Nevertheless, the theory of Rademacher type can be used to establish *global* results such as non-isomorphism of L_p-spaces: L_p is of type $\min\{2, p\}$, and so the isomorphic invariance of Rademacher type implies that L_p and L_q cannot be isomorphic whenever $1 \leq p, q \leq 2$ and $p \neq q$. As such local-to-global results abound in the theory of Banach spaces, it is of interest to discern Banach spaces with the "same" finite-dimensional subspaces.

Since all finite-dimensional Banach spaces of the same dimension are isomorphic to one another, we need a quantitative measure of how isomorphic two spaces are in order to devise a meaningful notion of sameness in this context. To this end, we introduce the following:

Definition 7.3. The *Banach–Mazur distance* between two isomorphic Banach spaces X and Y is given by

$$d(X, Y) = \inf\{\|T\|\|T^{-1}\| : T \in GL(X, Y)\},$$

where $GL(X, Y)$ is the collection of all isomorphisms of Banach spaces from X to Y.

With this notion, it now makes sense to talk about two Banach spaces whose finite-dimensional subspaces are "uniformly close to each other".

Definition 7.4 (James, 1972; [Jam72])**.** An infinite-dimensional Banach space X is *crudely finitely representable with constant C* in another infinite-dimensional Banach space Y if there exists a constant $C > 1$ such that each finite-dimensional subspace E of X admits a finite-dimensional subspace F of Y such that the Banach-Mazur distance $d(E, F)$ is bounded above by C.

Surprisingly, it is possible for non-isomorphic Banach spaces to have the "same" finite-dimensional subspaces:

Theorem 7.5 (Ribe rigidity theorem, 1976; [Rib76])**.** *If X and Y are uniformly homeomorphic Banach spaces, viz., there exists a uniformly continuous function $f : X \to Y$ whose inverse is also uniformly continuous, then X is crudely finitely representable in Y, and vice versa.*

In light of the Ribe rigidity theorem, we might reasonably wonder if we could forget the linear structure and develop a theory of type for metric spaces. To purse this, we must make sense of the notion of randomized signs in the absence of a linear structure. Let us first make the following definitions:

Definition 7.6. Let (M, d_M) be a metric space and $\mathbb{Z}_2^N$ the N-fold product of the cyclic group of order two with the counting measure, normalized to be of total measure 1. A *geometric N-cube* in (M, d_M) is the image of an injection $\mathbb{Z}_2^N \hookrightarrow M$.

Definition 7.7. Take the additive-group parametrization $\{0,1\}^N$ of $\mathbb{Z}_2^N$. The *l_p N-cube* is the geometric N-cube given by the map $f : \mathbb{Z}_2^N \to l_p$ that sends $(x_1, \ldots, x_N)$ to $(x_n)_{n=1}^N$, where $x_n = 0$ for all $n > N$. In other words, the l_p N-cube is $\{0,1\}^N$ with the metric

$$d(x, y) = \left(\sum_{n=1}^{N} |x_n - y_n|^p \right)^{1/p}.$$

The l_1 N-cube is called the *Hamming N-cube*, or the *N-dimensional Hamming cube.*

Taking the multiplicative-group parametrization $\{-1,1\}^N$ of $\mathbb{Z}_2^N$, we shall write $\{x_\varepsilon\}_{\varepsilon\in\{-1,1\}^N}$ to denote a geometric N-cube in (M,d_M). We define a *diagonal* of $\{x_\varepsilon\}_{\varepsilon\in\{-1,1\}^N}$ to be an unordered pair $\{x_\varepsilon, x_{-\varepsilon}\}$, and an *edge* of $\{x_\varepsilon\}_{\varepsilon\in\{-1,1\}^N}$ to be an unordered pair $\{x_\varepsilon, x_{\varepsilon'}\}$, where

$$\varepsilon' = \begin{cases} \varepsilon_n & \text{if } n \neq n_0; \\ -\varepsilon_n & \text{if } n = n_0; \end{cases} \tag{7.8}$$

for some fixed $1 \leq n_0 \leq N$. Given an injection $f : \{-1,1\}^N \to f : \{-1,1\}^N$ that gives rise to a geometric cube $\{x_\varepsilon\}_{\varepsilon\in\{-1,1\}^N}$, we write

$$\begin{aligned} \operatorname{diag}_f(\varepsilon) &= d(f(\varepsilon), f(-\varepsilon)) &= d(x_\varepsilon, x_{-\varepsilon}); \\ \operatorname{edge}_f(\varepsilon, n_0) &= d(f(\varepsilon), f(\varepsilon')) &= d(x_\varepsilon, x_{\varepsilon'}), \end{aligned} \tag{7.9}$$

where ε' is defined as in (7.8).

We now derive a metric analogue of Rademacher type. Let X be a Banach space of Rademacher type p. If $x_1, \ldots, x_N$ are distinct vetors in X, then the map

$$f(\varepsilon) = \sum_{n=1}^N \varepsilon_n x_n \tag{7.10}$$

is an injection from $\mathbb{Z}_2^N$ into X. Note that

$$\operatorname{diag}_f(\varepsilon) = \left\| \frac{f(\varepsilon) - f(-\varepsilon)}{2} \right\|,$$
$$\operatorname{edge}_f(\varepsilon, n) = \|\partial_n f(\varepsilon)\|,$$

where

$$\partial_n f(\varepsilon) = \frac{f(\varepsilon) - f(\varepsilon_1, \ldots, \varepsilon_{n-1}, -\varepsilon_n, \varepsilon_{n+1}, \ldots, \varepsilon_N)}{2}.$$

Therefore, we see that

$$\begin{aligned} \left\| \sum_{n=1}^N \varepsilon_n x_n \right\| &= \frac{1}{2}\|f(\varepsilon) - f(-\varepsilon)\| &= \frac{1}{2}\operatorname{diag}_f(\varepsilon) \\ \left(\sum_{n=1}^N \|x_n\|^p \right)^{\frac{1}{p}} &= \left(\sum_{n=1}^N \|\partial_n f(\varepsilon)\|^p \right)^{\frac{1}{p}} &= \frac{1}{2}\left(\sum_{n=1}^N \operatorname{edge}_f(\varepsilon, n)^p \right)^{\frac{1}{p}} \end{aligned}$$

for each $\varepsilon \in \mathbb{Z}_2^N$, whence we can rewrite (7.2) as

$$\left(\mathbb{E}_\varepsilon \left[\operatorname{diag}_f(\varepsilon)^p\right]\right)^{1/p} \leq C_p \left(\mathbb{E}_\varepsilon \left[\sum_{n=1}^N \operatorname{edge}_f(\varepsilon, n)^p \right] \right)^{1/p},$$

where $\mathbb{E}_\varepsilon$ is the expectation with respect to $\varepsilon \in \mathbb{Z}_2^N$. In light of this observation, we make the following definition:

Definition 7.11 (Enflo, 1969; [BMW86])**.** A metric space (M, d_M) is of *Enflo type* p for some $1 \leq p \leq 2$ if there exists a constant C_p, depending only on p, such that the inequality

$$\left(\mathbb{E}_\varepsilon \left[\operatorname{diag}_f(\varepsilon)^p\right]\right)^{1/p} \leq C_p \left(\mathbb{E}_\varepsilon \left[\sum_{n=1}^N \operatorname{edge}_f(\varepsilon, n)^p \right] \right)^{1/p}$$

holds for each geometric cube $f : \{-1,1\}^N \to M$ and every finite sequence $(x_n)_{n=1}^N$ in M, where $\mathbb{E}_\varepsilon$ is the expectation with respect to $\varepsilon \in \mathbb{Z}_2^N$. Here $\operatorname{diag}_f(\varepsilon)$ and $\operatorname{edge}_f(\varepsilon, n)$ are defined as in (7.9).

Example 7.12. The Hamming N-cube $\{-1,1\}^N$ is of Enflo type 2. To see this, we call upon Fourier analysis on the Boolean hypercube, discussed in Section 6 of the Appendix.

Fix $M \in \mathbb{N}$ and take a geometric cube $f : \{-1,1\}^M \to \{-1,1\}^N$ in the Hamming cube. Taking the Fourier–Walsh expansion of $\operatorname{diag}_f(\varepsilon)$, we see that

$$\begin{aligned} \operatorname{diag}_f(\varepsilon) = \sum_{S\subseteq[M]} \hat{f}(S)(\chi_S(\varepsilon) - \chi_S(-\varepsilon)) &= \sum_{S\subseteq[M]} \hat{f}(S)(1-(-1)^{|S|})\chi_S(\varepsilon) \\ &= \sum_{\substack{S\subseteq[M] \\ |S| \text{ odd}}} 2\hat{f}(S)\chi_S(\varepsilon). \end{aligned}$$

By uniqueness, the last expression is the Fourier–Walsh expansion of $\operatorname{diag}_f(\varepsilon)$. Parseval's identity implies that

$$\mathbb{E}_\varepsilon\left[\operatorname{diag}_f(\varepsilon)^2\right] = \sum_{\substack{S\subseteq[M] \\ |S| \text{ odd}}} 4\hat{f}(S)^2. \tag{7.13}$$

Let us now study $\operatorname{edge}_f(\varepsilon, m)$. To this end, we let

$$D_m x = (x_1, \ldots, -x_m, \ldots, x_M)$$

for each $1 \le m \le M$. Since

$$\chi_S(x) - \chi_S(D_m x) = \begin{cases} 2\chi_S(x) & \text{if } x \in S, \\ 0 & \text{if } x \notin S, \end{cases}$$

the Fourier–Walsh expansion of $\operatorname{edge}_f(\varepsilon, m)$ is

$$\operatorname{edge}_f(\varepsilon, m) = \sum_{S\subseteq[M]} \hat{f}(S)(\chi_S(x) - \chi_S(D_m x)) = \sum_{\substack{S\subseteq[M] \\ S\ni m}} 2\hat{f}(S)\chi_S(x).$$

By uniqueness, the last expression is the Fourier–Walsh expansion of $\operatorname{edge}_f(\varepsilon, m)$. Parseval's identity shows that

$$\frac{1}{2^M} \sum_{\varepsilon\in\mathbb{Z}_2^M} \operatorname{edge}_f(\varepsilon, m)^2 = \sum_{\substack{S\subseteq[M] \\ S\ni m}} 4\hat{f}(S)^2.$$

This implies that

$$\begin{aligned} \mathbb{E}_\varepsilon\left[\sum_{m=1}^M \operatorname{edge}_f(\varepsilon, m)^2\right] &= \frac{1}{2^M} \sum_{\varepsilon\in\mathbb{Z}_2^M} \sum_{m=1}^M \operatorname{edge}_f(\varepsilon, m)^2 \\ = \sum_{m=1}^M \left(\frac{1}{2^M} \sum_{\varepsilon\in\mathbb{Z}_2^M} \operatorname{edge}_f(\varepsilon, m)^2\right) &= \sum_{m=1}^M \sum_{\substack{S\subseteq[M] \\ S\ni m}} 4\hat{f}(S)^2 \\ = \sum_{S\subseteq[M]} 4|S|\hat{f}(S)^2 &\ge \sum_{S\subseteq[M]} 4\hat{f}(S)^2, \end{aligned}$$

as $|S| \geq 1$ for all $S \subseteq [M]$. It now follows from (7.13) that

$$\mathbb{E}_\varepsilon\left[\sum_{m=1}^{M} \operatorname{edge}_f(\varepsilon, m)^2\right] \geq \mathbb{E}_\varepsilon\left[\operatorname{diag}_f(\varepsilon)^2\right].$$

Since M was aribtrary, we conclude that the Hamming cube is of type 2 with constant at most 1. □

References

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov, *A course in metric geometry*, American Mathematical Society, 2001.

[BC02] Károly Bezdek and Robert Connelly, *Pushing disks apart – the kneser–poulsen conjecture in the plane*, Journal für die reine und angewandte Mathematik **2002** (2002), no. 553, 221–236.

[Bez10] Károly Bezdek, *Classical topics in discrete geometry*, Springer, 2010.

[BJ71] J. L. Bell and F. Jellett, *On the relationship between the boolean prime ideal theorem and two principles in functional analysis*, Bulletin de l'Académie Polonaise des Sciences. Série des Sciences Mathématiques, Astronomiques et Physiques **19** (1971), 191–194.

[BL00] Yoav Benyamini and Joram Lindenstrauss, *Geometric nonlinear functional analysis*, vol. 1, American Mathematical Society, 2000.

[BMW86] Jean Bourgain, Vitali D. Milman, and Haim J. Wolfson, *On type of metric spaces*, Transactions of the American Mathematical Society **294** (1986), no. 1, 295–317.

[Die10] Reinhard Diestel, *Graph theory*, Springer, 2010.

[Fol95] Gerald B. Folland, *A course in abstract harmonic analysis*, CRC Press, 1995.

[Fre] David H. Fremlin, *Kirszbraun's theorem*, http://www.essex.ac.uk/maths/people/fremlin/n11706.pdf.

[GZ68] F. Alberto Grünbaum and Eduardo Héctor Zarantonello, *On the extension of uniformly continuous mappings*, The Michigan Mathematics Journal **15** (1968), no. 1, 65–74.

[Jam72] Robert C. James, *Super-reflexive banach spaces*, Canadian Journal of Mathematics **24** (1972), 896–904.

[JLS86] William B. Johnson, Joram Lindenstrauss, and Gideon Schechtman, *Extension of lipschitz maps into banach spaces*, Israel Journal of Mathematics **54** (1986), no. 2, 129–138.

[Kir34] Mojżesz David Kirszbraun, *Über die zusammenziehende und lipschitzsche transformationen*, Fundamenta Mathematicae **22** (1934), no. 1, 77–108.

[Kne55] Von Martin Kneser, *Einige bemerkungen über das minkowskische flächenmass*, Archiv der Mathematik **6** (1955), 382–390.

[KW91] Victor Klee and Stan Wagon, *Old and new unsolved problems in plane geometry and number theory*, Dolciani Mathematical Expositions, Mathematical Association of America, 1991.

[Nao12] Assaf Naor, *An introduction to the ribe program*, Japanese Journal of Mathematics **7** (2012), 167–233.

[O'D] Ryan O'Donnell, *Analysis of boolean functions: Fall 2012 course at carnegie mellon*, http://www.contrib.andrew.cmu.edu/~ryanod/.

[Pou54] E. Thue Poulsen, *Problem 10*, Mathematica Scandinavica **2** (1954), 346.

[Rib76] Martin Ribe, *On uniformly homeomorphic normed spaces*, Arkiv För Matematik **14** (1976), 237–244.