Electronic Journal of Probability

Vol. 12 (2007), Paper no. 38, pages 1052–1071.

Journal URL
http://www.math.washington.edu/~ejpecp/

Interpolation of Random Hyperplanes*

Ery Arias-Castro†
Department of Mathematics
University of California, San Diego
9500 Gilman Drive, La Jolla CA 92093-0112, USA
mailto:eariasca@math.ucsd.edu
http://math.ucsd.edu/~eariasca/

Abstract

Let $\{(Z_i, W_i) : i = 1, \dots, n\}$ be uniformly distributed in $[0,1]^d \times \mathbb{G}(k,d)$, where $\mathbb{G}(k,d)$ denotes the space of k-dimensional linear subspaces of $\mathbb{R}^d$. For a differentiable function $f : [0,1]^k \to [0,1]^d$, we say that f interpolates $(z,w) \in [0,1]^d \times \mathbb{G}(k,d)$ if there exists $x \in [0,1]^k$ such that $f(x) = z$ and $\vec{f}(x) = w$, where $\vec{f}(x)$ denotes the tangent space at x defined by f. For a smoothness class $\mathcal{F}$ of Hölder type, we obtain probability bounds on the maximum number of points a function $f \in \mathcal{F}$ interpolates.

Key words: Grassmann Manifold, Haar Measure, Pattern Recognition, Kolmogorov Entropy.

AMS 2000 Subject Classification: Primary 60D05; Secondary: 62G10.

Submitted to EJP on September 13, 2006, final version accepted July 25, 2007.

*This work was partially supported by NSF grant DMS-0603890. The author was at the Mathematical Sciences Research Institute while preparing the first draft

†The author would like to thank Emmanuel Candès, David Donoho, Bruce Driver, Bo'az Klartag and Allen Knutson for helpful discussions and references.

1 Introduction

This paper is motivated by experiments in the field of Psychophysics (6) that study the ability of the Human Visual System at detecting curvilinear features in background clutter. In these experiments, human subjects are shown an image consisting of oriented small segments of same length dispersed in a square, such as in Figure 1.

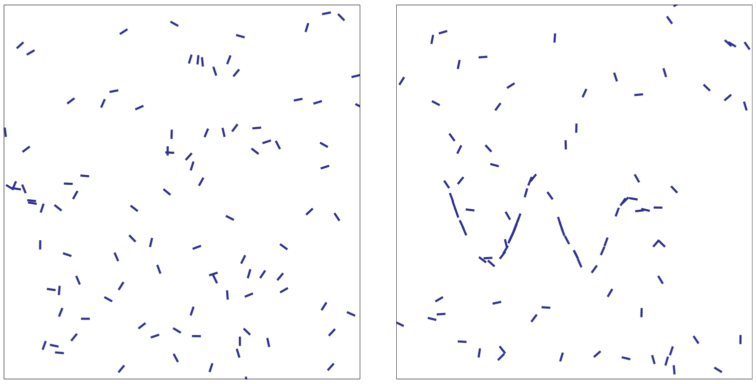

Figure 1: In Panel (a) we observe a realization under the null hypothesis ($n = 100$). In Panel (b) we observe a realization under the alternative hypothesis ($n = 100, n_1 = 40$).

The locations and orientations of these segments are either purely random (panel (a)) or a curve is actually "hidden" among purely random clutter, which here means that a curve was used to simulate a fraction of the segments by randomly choosing segments that are tangent to the curve at their midpoint (panel (b)).

From a Statistics viewpoint, this detection task, that human subjects are asked to perform, can be formalized into a hypothesis testing problem.

We say that a curve $\gamma \subset [0,1]^2$, parametrized by arclength, interpolates $(z,w) \in [0,1]^2 \times \mathbb{S}^1$ if there is x such that $\gamma(x) = z$ and $\dot{\gamma}(x) = w$, where $\mathbb{S}^1$ denotes the unit circle and $\dot{\gamma}(x)$ the derivative of γ at x.

We observe n segments of fixed length dispersed in the unit square.

- Under the *null hypothesis*, the segments have locations and orientations sampled uniformly at random in $[0,1]^2 \times \mathbb{S}^1$.
- Under the *(composite) alternative hypothesis*, the segments are as above except for n_1 of them that are chosen among those that a fixed curve γ interpolates. The curve γ is unknown but restricted to belong to some known class Γ.

Note that we do not specify the distribution of the segments tangent to the curve.

For $\gamma \in \Gamma$, define

$$N_n^{\rightarrow}(\gamma) = \#\{i = 1, \dots, n : \gamma \text{ interpolates } (Z_i, W_i)\},$$

and, with some abuse of notation,

$$N_n^{\rightarrow}(\Gamma) = \max_{\gamma \in \Gamma} N_n^{\rightarrow}(\gamma).$$

In (2), the test that rejects for large $N_n^{\rightarrow}(\Gamma)$ was analyzed for Γ the class of curves in the unit square with length and curvature bounded by some constant $c > 0$. In particular, it was shown that, under the null hypothesis, for some constants $0 < A < B < \infty$,

$$\mathbf{P}\left\{A\ n^{1/4} \leq N_n^{\rightarrow}(\Gamma) \leq B\ n^{1/4}\right\} \to 1, \quad n \to \infty.$$

Note that the upper bound implies that this test is powerful when $n_1 \geq Bn^{1/4}$.

A note on computing. The problem of computing $N_n^{\rightarrow}(\Gamma)$ exactly can be reframed into the *Reward-Budget Problem* of (5), the same problem appearing with different names elsewhere, e.g. (3; 4). Problems of this flavor are all NP-hard, though in theory there are algorithms that run in polynomial time that compute good approximations (5).

In contrast, when Γ is the set of graphs of functions of the first coordinate (say), with first and second derivatives bounded by some constants c_1 and c_2, a Dynamic Programming algorithm is available – see (8).

In this paper, we generalize this setting to higher dimensions. Let $\mathbb{G}(k,d)$ be the set of k-dimensional linear subspaces in $\mathbb{R}^d$. To $\mathbb{G}(k,d)$ we associate its uniform measure λ, which is the only invariant probability measure on $\mathbb{G}(k,d)$ that is invariant under the action of the orthogonal group $\mathbb{O}(d)$ – see (10), Section 1.

For a function $f : [0,1]^k \to [0,1]^d$ differentiable at x, let

$$\vec{f}(x) = \operatorname{span}\left\{\partial_s f(x) : s = 1, \dots, k\right\}.$$

A function $f : [0,1]^k \to [0,1]^d$ is said to interpolate $(z, w) \in [0,1]^d \times \mathbb{G}(k,d)$ if there exists $x \in [0,1]^k$ such that $f(x) = z$ and $\vec{f}(x) = w$.

We consider the following hypothesis testing problem. We observe

$$\{(Z_i, W_i) : i = 1, \dots, n\} \subset [0,1]^d \times \mathbb{G}(k,d).$$

- Under the *null hypothesis*, $\{(Z_i, W_i) : i = 1, \dots, n\}$ are independent and identically uniformly distributed in $[0,1]^d \times \mathbb{G}(k,d)$.
- Under the *(composite) alternative hypothesis*, $\{(Z_i, W_i) : i = 1, \dots, n\}$ are as above except for n_1 of them that are chosen among those that a fixed function f interpolates. The function f is unknown but restricted to belong to some known class $\mathcal{F}$.

Before specifying $\mathcal{F}$, we introduce some notation. For a vector $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, the supnorm is defined as $\|x\|_\infty = \max\{|x_i| : i = 1, \dots, d\}$. For a function $f : \Omega \subset \mathbb{R}^k \to \mathbb{R}^d$, $\|f\|_\infty = \sup_{x \in \Omega} \|f(x)\|_\infty$. The Euclidean inner product and the corresponding norm are denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ respectively. The angle $\angle(H,K) \in [0,\pi]$ between two linear subspaces $H, K \subset \mathbb{R}^d$, with $1 \leq \dim H \leq \dim K$, is defined by

$$\angle(H,K) = \max_{u \in H} \min_{v \in K} \arccos\left(\frac{\langle u, v \rangle}{\|u\|\|v\|}\right).$$

This corresponds to the largest canonical angle as defined in (7) and constitutes a metric on $\mathbb{G}(k,d)$ – see also (1) for a related study of the largest canonical angle between two subspaces uniformly distributed in $\mathbb{G}(k,d)$.

The class $\mathcal{F}$, parametrized by $\beta \geq 1$, is defined as the set of twice differentiable, one-to-one functions $f : [0,1]^k \to [0,1]^d$ with the following additional properties:

- For all $s = 1, \ldots, k$, $1/\beta \leq \|\partial_s f(x)\|_\infty \leq \beta$ for all $x \in [0,1]^k$;
- For all $s, t = 1, \ldots, k$, $\|\partial_{st} f(x)\|_\infty \leq \beta$ for all $x \in [0,1]^k$;
- For all $s = 1, \ldots, k$ and $x \in [0,1]^k$,

$$\angle(\partial_s f(x), \mathrm{span}\{\partial_t f(x) : t \neq s\}) \geq \frac{1}{2\beta(d-k)},$$

  which is void if $k = 1$. (In this paper, we identify a non-zero vector with the one dimensional linear subspace it generates.)

The last condition and the constraint $\beta \geq 1$ ensure that $\mathcal{F}$ contains graphs of the form $x \to (x, g(x))$, where $g : [0,1]^k \to [0,1]^{d-k}$ satisfies the first two conditions – see Lemma 6.1.
Define

$$N_n^{\rightarrow}(f) = \#\{i = 1, \ldots, n : f \text{ interpolates } (Z_i, W_i)\},$$

and, with some abuse of notation,

$$N_n^{\rightarrow}(\mathcal{F}) = \max_{f \in \mathcal{F}} N_n^{\rightarrow}(f).$$

Let

$$\vec{\rho} = \frac{k}{k + (d-k)(k+2)}.$$

Theorem 1.1. *There are finite, positive constants A and B depending only on k, d, β such that, under the null hypothesis,*

$$\mathbf{P}\left\{A\, n^{\vec{\rho}} \leq N_n^{\rightarrow}(\mathcal{F}) \leq B\, n^{\vec{\rho}}\right\} \to 1, \quad n \to \infty.$$

As before, this implies that the test that rejects for large values of $N_n^{\rightarrow}(\mathcal{F})$ is powerful when $n_1 > B n^{\vec{\rho}}$.

The remaining of the paper is organized as follows. In Section 2 we introduce a related, yet different hypothesis testing problem. In Section 3 and Section 4, we prove results announced in Section 2. In Section 5 and Section 6, we follow the arguments in Section 3 and Section 4 to prove Theorem 1.1. Some intermediary lemmas are proved in the Appendix.

2 Another Hypotheses Testing Problem

We introduce another hypothesis testing problem as a stepping stone towards proving Theorem 1.1 and also for its own sake.

Let $\alpha > 1$, $\beta > 0$ and define $r = \lfloor \alpha \rfloor = \max\{m \in \mathbb{N} : m < \alpha\}$. (In this paper, we include 0 in $\mathbb{N}$.) Define the Hölder class $\mathrm{H}^{k,d}(\alpha, \beta)$ to be the set of functions $f : [0,1]^k \to [0,1]^d$, with $f = (f_1, \dots, f_d)$ such that, for all $\mathbf{s} = (s_1, \dots, s_k) \in \mathbb{N}^k$ with $|\mathbf{s}| = s_1 + \cdots + s_k \leq r$,

$$\|f^{(\mathbf{s})}(x)\|_\infty \leq \beta \quad \forall x \in [0,1]^k;$$

and, for all $\mathbf{s} \in \mathbb{N}^k$ with $|\mathbf{s}| = r$,

$$\|f^{(\mathbf{s})}(x) - f^{(\mathbf{s})}(y)\|_\infty \leq \beta \|x - y\|_\infty^{\alpha - r} \quad \forall x, y \in [0,1]^k,$$

where $f^{(\mathbf{s})} = (\partial_{s_1 \cdots s_k} f_1, \dots, \partial_{s_1 \cdots s_k} f_d)$. When there is no possible confusion, we use the notation $\mathcal{H} = \mathrm{H}^{k,d-k}(\alpha, \beta)$.

Fix r_0 an integer such that $1 \leq r_0 \leq r$. Let

$$\mathbf{S} = \{\mathbf{s} \in \mathbb{N}^k : |\mathbf{s}| \leq r_0\},$$

with cardinality

$$|\mathbf{S}| = \sum_{s=0}^{r_0} \binom{s + k - 1}{k - 1}.$$

We denote by $y^{\mathbf{S}}$ a vector in $\mathbb{R}^{(d-k)|\mathbf{S}|}$ and by $f^{(\mathbf{S})}(x)$ the vector $(f^{(\mathbf{s})}(x) : \mathbf{s} \in \mathbf{S})$. A function $f \in \mathcal{H}$ is said to interpolate $(x, y^{\mathbf{S}}) \in [0,1]^k \times \mathbb{R}^{(d-k)|\mathbf{S}|}$ if $f^{(\mathbf{S})}(x) = y^{\mathbf{S}}$.

Consider the following hypothesis testing problem. We observe

$$\{(X_i, Y_i^{\mathbf{S}})) : i = 1, \dots, n\} \subset [0,1]^k \times \mathbb{R}^{(d-k)|\mathbf{S}|}.$$

- Under the *null hypothesis*, $\{(X_i, Y_i^{\mathbf{S}}) : i = 1, \dots, n\}$ are independent and identically uniformly distributed in $[0,1]^k \times [0,1]^{d-k} \times [-\beta, \beta]^{(d-k)(|\mathbf{S}|-1)}$.
- Under the *(composite) alternative hypothesis*, $\{(X_i, Y_i^{\mathbf{S}}) : i = 1, \dots, n\}$ are as above except for n_1 of them that are chosen among those that a fixed function f interpolates. The function f is unknown but restricted to belong to $\mathcal{H}$.

Figure 2 shows an example, with $d = 2$ and $r_0 = 1$.

Define

$$N_n^{(r_0)}(f) = \#\{i = 1, \dots, n : f \text{ interpolates } (X_i, Y_i^{\mathbf{S}})\},$$

and, with some abuse of notation,

$$N_n^{(r_0)}(\mathcal{H}) = \max_{f \in \mathcal{H}} N^{(r_0)}(f).$$

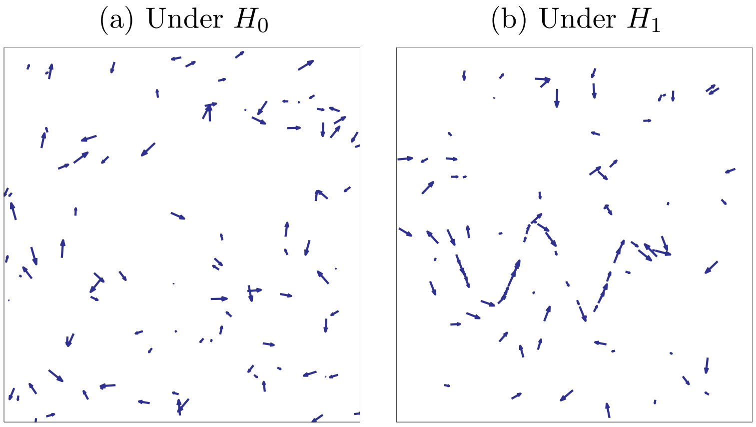

Figure 2: In Panel (a) we observe a realization under the null hypothesis ($n = 100$). In Panel (b) we observe a realization under the alternative hypothesis ($n = 100, n_1 = 40$).

Let

$$\rho(r_0) = \frac{k}{k + \alpha(d-k)w} = \frac{1}{1 + \alpha(d/k - 1)w} \ ,$$

where

$$w = \sum_{s=0}^{r_0} (1 - s/\alpha) \left(\begin{array}{c} s+k-1 \\ k-1 \end{array} \right).$$

Theorem 2.1. *There are finite, positive constants A and B depending only on k, d, α, β, r_0 such that, under the null hypothesis,*

$$\mathbf{P}\left\{A \ n^{\rho(r_0)} \le N_n^{(r_0)}(\mathcal{H}) \le B \ n^{\rho(r_0)}\right\} \to 1, \quad n \to \infty.$$

As before, this implies that the test that rejects for large values of $N_n^{(r_0)}(\mathcal{H})$ is powerful when $n_1 > Bn^{\rho(r_0)}$.

Remark. For $\alpha = 2$ and $r_0 = 1$, $\rho(r_0) = \vec{\rho}$, meaning that $N_n^{(r_0)}(\mathcal{H})$ and $N_n^{\rightarrow}(\mathcal{F})$ are, in that case, of same order of magnitude with high probability. This will be used explicitly in Section 6 when proving the lower bound in Theorem 1.1.

3 Proof of the upper bound in Theorem 2.1

For $y_1^{\mathbf{S}}, y_2^{\mathbf{S}} \in \mathbb{R}^{(d-k)|\mathbf{S}|}$, define the discrepancy

$$\Phi(y_1^{\mathbf{S}}, y_2^{\mathbf{S}}) = \max_{\mathbf{s} \in \mathbf{S}} \|y_1^{\mathbf{s}} - y_2^{\mathbf{s}}\|_\infty^{\alpha/(\alpha - |\mathbf{s}|)}.$$

The discrepancy Φ induces a discrepancy on functions, namely

$$\Phi(f, g) = \max_{\mathbf{s} \in \mathbf{S}} \|f^{(\mathbf{s})} - g^{(\mathbf{s})}\|_\infty^{\alpha/(\alpha - |\mathbf{s}|)}.$$

The argument for proving the upper bound in Theorem 2.1 is based on coverings of $\mathcal{H}$ with respect to Φ. Let L_ε be the ε-covering number of $\mathcal{H}$ with respect to Φ.

Lemma 3.1. *There is a constant $c_1 = c_1(k, d, \alpha, \beta, r_0) > 0$ such that, for all $\varepsilon > 0$,*

$$\log L_\varepsilon \le c_1 \varepsilon^{-k/\alpha}.$$

Lemma 3.1 follows immediately from the proof of Theorem XIII in (9), Chapter "ε-entropy and ε-capacity of sets in functional spaces".

For a set $K \subset \mathbb{R}^k \times \mathbb{R}^{(d-k)|\mathbf{S}|}$ and $\varepsilon > 0$, we denote by $K^{\varepsilon,\Phi}$ the set of points $(x, y^{\mathbf{S}})$ such that there is $(x_1, y_1^{\mathbf{S}}) \in K$ with $x_1 = x$ and $\Phi(y_1^{\mathbf{S}}, y^{\mathbf{S}}) \le \varepsilon$.
For each ε, we select an ε-net $\{f_j : j = 1, \dots, L_\varepsilon\}$ of $\mathcal{H}$ for Φ. For $j = 1, \dots, L_\varepsilon$, we define

$$\begin{aligned} K_j &= \{(x, y^{\mathbf{S}}) \in \mathbb{R}^k \times \mathbb{R}^{(d-k)|\mathbf{S}|} : \Phi(y^{\mathbf{S}}, f_j^{(\mathbf{S})}(x)) \le \varepsilon\} \\ &= \mathrm{graph}^{\mathbf{S}}(f_j)^{\varepsilon,\Phi}, \end{aligned}$$

where, for $f \in \mathcal{H}$,

$$\mathrm{graph}^{\mathbf{S}}(f) = \{(x, f^{(\mathbf{S})}(x)) : x \in [0,1]^k\} \subset [0,1]^k \times \mathbb{R}^{(d-k)|\mathbf{S}|}.$$

We extend $N^{(r_0)}(\cdot)$ to subsets $K \subset \mathbb{R}^k \times \mathbb{R}^{(d-k)|\mathbf{S}|}$, by defining $N^{(r_0)}(K)$ to be the number of points $(X_i, Y_i^{\mathbf{S}})$ that belong to K.
Let

$$M_n(\varepsilon) = \max_{j=1,\dots,L_\varepsilon} N_n^{(r_0)}(K_j).$$

By definition, it is straightforward to see that

$$N_n^{(r_0)}(\mathcal{H}) \le M_n(\varepsilon),$$

for all $\varepsilon > 0$. We therefore focus on bounding $M_n(\varepsilon)$.

By Boole's inequality, we have

$$\mathbf{P}\{M_n(\varepsilon) > b\} \le L_\varepsilon \cdot \max_{j=1,\dots,L_\varepsilon} \mathbf{P}\left\{N_n^{(r_0)}(K_j) > b\right\}.$$

Moreover, we know that, for any set $K \subset \mathbb{R}^k \times \mathbb{R}^{(d-k)|\mathbf{S}|}$,

$$N_n^{(r_0)}(K) \backsim \mathrm{Bin}(n, \mu(K)),$$

where μ is the uniform measure on $[0,1]^k \times [0,1]^{d-k} \times [-\beta, \beta]^{(d-k)(|\mathbf{S}|-1)}$.
Hence,

$$\begin{aligned} \mathbf{P}\{M_n(\varepsilon) > b\} &\le L_\varepsilon \cdot \max_{j=1,\dots,L_\varepsilon} \mathbf{P}\{\mathrm{Bin}(n, \mu(K_j)) > b\} \\ &= L_\varepsilon \cdot \mathbf{P}\left\{\mathrm{Bin}(n, \max_j \mu(K_j)) > b\right\}. \end{aligned}$$

Lemma 3.2. *There is $c_2 > 0$ such that, for all $f \in \mathcal{H}$ and all $\varepsilon > 0$,*

$$\mu\left(\mathrm{graph}^{\mathbf{S}}(f)^{\varepsilon,\Phi}\right) \le c_2 \varepsilon^{(d-k)w}.$$

Proof of Lemma 3.2. Start with

$$\mathrm{graph}^{\mathbf{S}}(f)^{\varepsilon,\Phi} \subset \cup_{x\in[0,1]^k} \{x\} \bigotimes_{\mathbf{s}\in\mathbf{S}} \mathbb{B}(f^{(\mathbf{s})}(x), \varepsilon^{1-|\mathbf{s}|/\alpha}),$$

where $\mathbb{B}(y,\eta)$ denotes the supnorm ball in $\mathbb{R}^{d-k}$ centered at y of radius η.
Hence, integrating over $x \in [0,1]^k$ last, we have

$$\mu(\mathrm{graph}^{\mathbf{S}}(f)^{\varepsilon,\Phi}) \le c_2 \prod_{\mathbf{s}\in\mathbf{S}} (\varepsilon^{1-|\mathbf{s}|/\alpha})^{d-k} = c_2\ \varepsilon^{(d-k)w}.$$

□

Using Lemma 3.2, we arrive at

$$\mathbf{P}\{M_n(\varepsilon) > b\} \le L_\varepsilon \cdot \mathbf{P}\left\{\mathrm{Bin}(n, c_2\varepsilon^{(d-k)w}) > b\right\}.$$

Lemma 3.3. *There is a constant $c > 0$ such that, for any n positive integer and $p \in (0,1/2)$, and for all $b > 2np$,*

$$\mathbf{P}\{\mathrm{Bin}(n,p) > b\} \le \exp(-c \cdot b).$$

Lemma 3.3 follows directly from standard large deviations bounds for binomial random variables – see (12), p. 440, Inequality 1, (ii).
We use Lemma 3.3 to obtain

$$\mathbf{P}\left\{\mathrm{Bin}(n, c_2\varepsilon^{(d-k)w}) > b\right\} \le \exp(-c\cdot b),$$

for all $b > 2nc_2\varepsilon^{(d-k)w}$.
Collecting terms, we arrive at the following inequality, valid for $B > 2c_2$,

$$\mathbf{P}\left\{M_n(\varepsilon) > B\ \varepsilon^{(d-k)w} n\right\} \le \exp\left(c_1\varepsilon^{-k/\alpha} - c_3\ B\ \varepsilon^{(d-k)w} n\right).$$

Choose $\varepsilon = n^{-\alpha/(k+\alpha(d-k)w)}$, so that $\varepsilon^{-k/\alpha} = \varepsilon^{(d-k)w} n = n^{\rho(r_0)}$. Then, the result above transforms into

$$\mathbf{P}\left\{M_n(\varepsilon) > B\ n^{\rho(r_0)}\right\} \to 0, \quad n \to \infty,$$

valid for $B > (c_1/c_3) \vee 2c_2$.

4 Proof of the lower bound in Theorem 2.1

We use the notations appearing in Section 4, except for the various constants which are refreshed in this section.
For each $\mathbf{s} \in \mathbf{S}$, take $\psi_{\mathbf{s}} : \mathbb{R}^k \to \mathbb{R}$ infinitely differentiable, supported in $[-1/2, 1/2]^k$, and satisfying $\psi_{\mathbf{s}}^{(\mathbf{t})}(0) = 0$ if $\mathbf{t} \in \mathbf{S}$ and $\mathbf{t} \ne \mathbf{s}$; and $\psi_{\mathbf{s}}^{(\mathbf{s})}(0) = 1$.
Let $c_1 \ge 1$ such that $c_1 \ge \|\psi_{\mathbf{s}}^{(\mathbf{t})}\|_\infty$, for all $\mathbf{s} \in \mathbf{S}$ and $\mathbf{t} \in \mathbb{N}^k$ with $|\mathbf{t}| \le r + 1$.

Again, choose $\varepsilon > 0$ such that $\varepsilon^{-k/\alpha} = \varepsilon^{(d-k)w} n$. Define $\varepsilon_{\mathbf{s}} = \varepsilon^{1-|\mathbf{s}|/\alpha}$ and, with $c_2 > 1$ to be determined later, let $\varepsilon' = (c_2\varepsilon)^{1/\alpha}$. We partition $[0,1]^k$ into hypercubes of sidelength ε', that we denote by I_m, where $m = (m_1, \ldots, m_k) \in ([0, 1/\varepsilon'] \cap \mathbb{N})^k$. For a given m, let

$$R_m = I_m \times [\varepsilon/2, \varepsilon]^{d-k} \times \prod_{\mathbf{s} \in \mathbf{S} \setminus \{\mathbf{0}\}} [0, \varepsilon_{\mathbf{s}}]^{d-k} \subset \mathbb{R}^k \times \mathbb{R}^{(d-k)|\mathbf{S}|}.$$

Denote by $\mathcal{M} = ([0, 1/\varepsilon'] \cap 2\mathbb{N})^k$ and let M be the (random) subset of $m \in \mathcal{M}$ such that there is $i(m) \in \{1, \ldots, n\}$ with $(X_{i(m)}, Y^{\mathbf{S}}_{i(m)}) \in R_m$.

Lemma 4.1. *There is a constant $A = A(k, d, \alpha, \beta, r_0) > 0$ such that*

$$\mathbf{P}\left\{|M| > A\ n^{\rho(r_0)}\right\} \to 1, \quad n \to \infty.$$

Lemma 4.1 is proved in Appendix A. The lower bound in Theorem 2.1 now follows if we are able to interpolate the points $\{(X_{i(m)}, Y^{\mathbf{S}}_{i(m)}) : m \in M\}$ with a function in $\mathcal{H}$.

Define, for each $j = 1, \ldots, d-k$,

$$h_j(x) = \sum_{m \in M} h_{j,m}(x),$$

where

$$h_{j,m}(x) = g_{j,m}\left(\frac{x - X_{i(m)}}{\varepsilon'}\right),$$

with

$$g_{j,m}(x) = Y^{\mathbf{0}}_{i(m),j}\ \psi_{\mathbf{0}}(x)\ \sum_{\mathbf{s} \in \mathbf{S}} (\varepsilon')^{|\mathbf{s}|}\ \frac{Y^{\mathbf{s}}_{i(m),j}}{Y^{\mathbf{0}}_{i(m),j}}\ \psi_{\mathbf{s}}(x).$$

Note that, for $\mathbf{t} \in \mathbb{N}^k$,

$$g^{(\mathbf{t})}_{j,m}(x) = Y^{\mathbf{0}}_{i(m),j} \sum_{\mathbf{0} \le \mathbf{t}' \le \mathbf{t}} \left(\begin{array}{c} \mathbf{t} \\ \mathbf{t}' \end{array}\right)\ \psi^{(\mathbf{t}')}_{\mathbf{0}}(x)\ \sum_{\mathbf{s} \in \mathbf{S}} (\varepsilon')^{|\mathbf{s}|}\ \frac{Y^{\mathbf{s}}_{i(m),j}}{Y^{\mathbf{0}}_{i(m),j}}\ \psi^{(\mathbf{t}-\mathbf{t}')}_{\mathbf{s}}(x),$$

where if $\mathbf{s} = (s_1, \ldots, s_k)$ and $\mathbf{t} = (t_1, \ldots, t_k)$,

$$\left(\begin{array}{c} \mathbf{t} \\ \mathbf{s} \end{array}\right) = \left(\begin{array}{c} t_1 \\ s_1 \end{array}\right) \cdots \left(\begin{array}{c} t_k \\ s_k \end{array}\right).$$

Hence, $g^{(\mathbf{s})}_{j,m}(0) = Y^{\mathbf{s}}_{i(m),j}$. This implies that $h^{(\mathbf{s})}_{j,m}(X_{i(m)}) = Y^{\mathbf{s}}_{i(m),j}$, and so $h^{(\mathbf{s})}_{j}(X_{i(m)}) = Y^{\mathbf{s}}_{i(m),j}$, since for each $j = 1, \ldots, d-k$, the $h_{j,m}$'s have non-overlapping supports. Therefore, if we let $h = (h_1, \ldots, h_{d-k})$, we have that h interpolates each point in $\{(X_{i(m)}, Y^{\mathbf{S}}_{i(m)}) : m \in M\}$.

Remains to prove that, for each $j = 1, \ldots, d-k$, $h_j \in \mathrm{H}^{k,1}(\alpha, \beta)$. Fix $j = 1, \ldots, d-k$. Again, because for every $x \in [0,1]^k$ there is at most one m such that $h_{j,m}(x) \neq 0$, we have for all $\mathbf{t} \in \mathbb{N}^k$,

$$\|h^{(\mathbf{t})}_j\|_\infty \le \max_{m \in M} \|h^{(\mathbf{t})}_{j,m}\|_\infty.$$

Fix $m \in M$. We have $\|h^{(\mathbf{t})}_{j,m}\|_\infty = \|g^{(\mathbf{t})}_{j,m}\|_\infty \ (\varepsilon')^{-|\mathbf{t}|}$, with

$$\|g^{(\mathbf{t})}_{j,m}\|_\infty \leq Y^{\mathbf{0}}_{i(m),j} \sum_{\mathbf{0}\leq\mathbf{t}'\leq\mathbf{t}} \begin{pmatrix} \mathbf{t} \\ \mathbf{t}' \end{pmatrix} \|\psi^{(\mathbf{t}')}_{\mathbf{0}}\|_\infty \sum_{\mathbf{s}\in\mathbf{S}} (\varepsilon')^{|\mathbf{s}|} \frac{Y^{\mathbf{s}}_{i(m),j}}{Y^{\mathbf{0}}_{i(m),j}} \|\psi^{(\mathbf{t}-\mathbf{t}')}_{\mathbf{s}}\|_\infty.$$

Since $Y^{\mathbf{0}}_{i(m),j} \leq \varepsilon$ and

$$0 \leq (\varepsilon')^{|\mathbf{s}|} \frac{Y^{\mathbf{s}}_{i(m),j}}{Y^{\mathbf{0}}_{i(m),j}} \leq 2c_2^{|\mathbf{s}|/\alpha} \leq 2c_2^{r/\alpha},$$

we have, for $|\mathbf{t}| \leq r+1$, $\|g^{(\mathbf{t})}_{j,m}\|_\infty \leq c_3 \ c_2^{r/\alpha} \ \varepsilon$, with $c_3 = c_3(\alpha, k) > 0$. In particular, c_3 does not depend on the choice of c_2. Choose $c_2 > 1$ such that $c_3 \ c_2^{r/\alpha - 1} \leq \beta$.
Hence, for all $\mathbf{t}$ such that $|\mathbf{t}| \leq r+1$,

$$\|h^{(\mathbf{t})}_{j,m}\|_\infty \leq c_3 \ c_2^{r/\alpha} \ \varepsilon \ (\varepsilon')^{-|\mathbf{t}|} = c_3 \ c_2^{(r-|\mathbf{t}|)/\alpha} \ \varepsilon^{1-|\mathbf{t}|/\alpha}.$$

This implies that, for ε small enough, h_j takes values in $[0,1]$ and

$$\|h^{(\mathbf{t})}_j\|_\infty \leq \beta,$$

for all $\mathbf{t}$ such that $|\mathbf{t}| \leq r$.
Remains to prove that, for all $\mathbf{t}$ such that $|\mathbf{t}| = r$,

$$|h^{(\mathbf{t})}_j(x') - h^{(\mathbf{t})}_j(x)| \leq \beta \|x' - x\|_\infty^{\alpha - r},$$

for all $x, x' \in [0,1]^k$.

- Suppose $\|x' - x\|_\infty > \varepsilon'$;

$$\begin{aligned} |h^{(\mathbf{t})}_j(x') - h^{(\mathbf{t})}_j(x)| &\leq \|h^{(\mathbf{t})}_j\|_\infty \\ &\leq c_3 \ c_2^{r/\alpha} \ \varepsilon(\varepsilon')^{-r} \\ &\leq c_3 \ c_2^{r/\alpha} \ \varepsilon(\varepsilon')^{-r}(\|x - x'\|_\infty/\varepsilon')^{\alpha - r} \\ &\leq c_3 \ c_2^{r/\alpha - 1} \ \|x - x'\|_\infty^{\alpha - r}. \end{aligned}$$

- Suppose $\|x - x'\|_\infty \leq \varepsilon'$ and let $\mathbf{t}_+ = (t_1 + 1, \ldots, t_k)$;

$$\begin{aligned} |h^{(\mathbf{t})}_j(x') - h^{(\mathbf{t})}_j(x)| &\leq \|h^{(\mathbf{t}_+)}_j\|_\infty \|x - x'\|_\infty \\ &\leq c_3 \ c_2^{r/\alpha} \ \varepsilon(\varepsilon')^{-(r+1)} \cdot (\varepsilon')^{1-(\alpha - r)} \|x - x'\|_\infty^{\alpha - r} \\ &= c_3 \ c_2^{r/\alpha - 1} \ \|x - x'\|_\infty^{\alpha - r}. \end{aligned}$$

Since we chose c_2 such that $c_3 \ c_2^{r/\alpha - 1} \leq \beta$, we are finished.

5 Proof of the upper bound in Theorem 1.1

Let Ψ be the discrepancy on $\mathbb{R}^d \times \mathbb{G}(k,d)$ defined by

$$\Psi((z,H),(z_1,H_1)) = \max\{\|z - z_1\|_\infty, \angle(H,H_1)^2\}.$$

Each $f \in \mathcal{F}$ is identified with $(f, \vec{f})$, viewed as a function on $[0,1]^k$ with values in $\mathbb{R}^d \times \mathbb{G}(k,d)$ – the first and third constraints on the derivatives of $f \in \mathcal{F}$ guarantee that $\vec{f}(x)$ is indeed a k-dimensional subspace of $\mathbb{R}^d$ for all $x \in [0,1]^k$. With this perspective, Ψ induces a discrepancy on $\mathcal{F}$. The proof is based on coverings of $\mathcal{F}$ with respect to that discrepancy – still denoted by Ψ.

It turns out that Ψ is dominated by the discrepancy Φ defined in Section 4, with $\alpha = 2$ and $r_0 = 1$. Indeed, we have the following.

Lemma 5.1. *There is a constant $c = c(k,d,\beta)$ such that, for any $f, g \in \mathcal{F}$ and $x \in [0,1]^k$,*

$$\angle\left(\vec{f}(x), \vec{g}(x)\right) \le c \max_{s=1,\dots,k} \|\partial_s f(x) - \partial_s g(x)\|_\infty.$$

To get Lemma 5.1, we apply Lemma D.1 in Appendix D with u_i (resp. v_i) defined as $\partial_i f(x)$ (resp. $\partial_i g(x)$) and $c_1 = 1/\beta$, $c_2 = \beta$, $c_3 = 1/(2\beta(d-k))$.

Therefore, the ε-covering number of $\mathcal{F}$ with respect to Ψ is bounded by the ε-covering number of $\mathcal{F}$ with respect to Φ, whose logarithm is of order $\varepsilon^{-k/2}$ – see Lemma 3.1, where d enters only in the constant.

Following the steps in Section 4, we only need to find an equivalent of Lemma 3.2, namely compute an upper bound on the measure of the ε-neighborhood of

$$\{(f(x), \vec{f}(x)) : x \in [0,1]^k\} \subset \mathbb{R}^d \times \mathbb{G}(k,d)$$

for the discrepancy Ψ, valid for all $f \in \mathcal{F}$. For $H \in \mathbb{G}(k,d)$, let

$$B(H,\varepsilon) = \{K \in \mathbb{G}(k,d) : \angle(H,K) \le \varepsilon\}.$$

As in the proof of Lemma 3.2, we are left with computing a upper bound on $\lambda(B(H,\varepsilon))$, which is independent of $H \in \mathbb{G}(k,d)$ since λ is invariant under the (transitive) action of the orthogonal group. (Remember that λ denotes the uniform measure on $\mathbb{G}(k,d)$.)

Lemma 5.2. *There is a constant $c = c(k,d)$ such that, for all $\varepsilon > 0$ and for all $H \in \mathbb{G}(k,d)$, $\lambda(B(H,\varepsilon)) \le c\, \varepsilon^{(d-k)k}$.*

Lemma 5.2 is a direct consequence of Lemma B.1 in Appendix B and the fact that $\lambda(\mathbb{G}(k,d)) = 1$.

6 Proof of the lower bound in Theorem 1.1

We show that the lower bound in Theorem 2.1 implies the lower bound in Theorem 1.1. We first start by showing that functions of the form $f(x) = (x, g(x))$, with $g \in \mathcal{H}$, belong to $\mathcal{F}$.

Lemma 6.1. *For all $g \in \mathcal{H}$, the function $f(x) = (x, g(x))$ belongs to $\mathcal{F}$.*

Lemma 6.1 is proved in Appendix C.

Let W be sampled uniformly at random in $\mathbb{G}(k,d)$. With probability one, there is a unique set of vectors in $\mathbb{R}^{d-k}$, $\{Y^{\mathbf{s}} : |\mathbf{s}| = 1\}$, such that $W = \operatorname{span}\{(\mathbf{s}, Y^{\mathbf{s}}) : |\mathbf{s}| = 1\}$. Indeed, W has the same distribution as $\operatorname{span}\{w_1, \dots, w_k\}$, where $w_1, \dots, w_k$ are i.i.d. uniformly distributed on the unit sphere of $\mathbb{R}^d$ and therefore, with probability one, $\langle w_i, e_i \rangle \neq 0$ for all $i = 1, \dots, k$, $\{e_i : i = 1, \dots, d\}$ being the canonical basis of $\mathbb{R}^d$. The uniqueness comes from the fact that a subspace of the form $\operatorname{span}\{(\mathbf{s}, Y^{\mathbf{s}}) : |\mathbf{s}| = 1\}$ does not contain a vector of the form $(0, Y)$, with $Y \in \mathbb{R}^{d-k} \setminus \{0\}$, so it does not contain two distinct vectors of the form $(\mathbf{s}, Y_1)$ and $(\mathbf{s}, Y_2)$.

Through the map κ that associates W to $\{Y^{\mathbf{s}} : |\mathbf{s}| = 1\}$, the uniform measure on $\mathbb{G}(k,d)$ induces a probability measure ν on $\mathbb{R}^{(d-k)k}$. If we observe $\{(Z_i, W_i) : i = 1, \dots, n\}$, we let

$$Z_i = (X_i, Y_i) \in \mathbb{R}^k \times \mathbb{R}^{d-k} \quad \text{and} \quad W_i = \operatorname{span}\{(\mathbf{s}, Y_i^{\mathbf{s}}) : |\mathbf{s}| = 1\}.$$

With here $\mathbf{S} = \{\mathbf{s} \in \mathbb{N}^k : |\mathbf{s}| \leq 1\}$, we thus obtain $\{(X_i, Y_i^{\mathbf{S}}) : i = 1, \dots, n\}$, independent and with common distribution $\Lambda_k \otimes \Lambda_{d-k} \otimes \nu \equiv \Lambda_d \otimes \nu$, where Λ_ℓ is the uniform measure on $[0,1]^\ell$. Note that, if $g \in \mathcal{H}$ interpolates $\{(X_i, Y_i^{\mathbf{S}}) : i = 1, \dots, n\}$, then f defined by $f(x) = (x, g(x))$ belongs to $\mathcal{F}$ by Lemma 6.1 and interpolates $\{(Z_i, W_i) : i = 1, \dots, n\}$. With $\Lambda_d \otimes \nu$ playing the role of μ, the uniform measure on $[0,1]^k \times [0,1]^{d-k} \times [-\beta, \beta]^{(d-k)(|\mathbf{S}|-1)}$, the present setting parallels the situation in Section 4. Following the arguments given there, we are only left with obtaining the equivalent of Lemma 4.1. Looking at the proof of Lemma 4.1 in Section A, all we need is a lower bound of the form

$$\Lambda_d \otimes \nu(R_{\mathbf{0}}) \geq c\ \mu(R_{\mathbf{0}}) = c\ (\varepsilon')^k \varepsilon^{(d-k)w} = c\ \varepsilon^{k/2 + (d-k)(1+k/2)}.$$

(Here $\alpha = 2$ and $w = 1 + k/2$.) Because

$$\Lambda_d \otimes \nu(R_{\mathbf{0}}) = c\ \varepsilon^{k/2 + (d-k)} \cdot \nu\left([0, \varepsilon^{1/2}]^{(d-k)k}\right),$$

the following lemma provides what we need.

Lemma 6.2. *There is a constant $c = c(k,d) > 0$ such that, for all $\varepsilon > 0$ small enough,*

$$\nu\left([0,\varepsilon]^{(d-k)k}\right) > c\ \varepsilon^{(d-k)k}.$$

To prove Lemma 6.2, we first show that for some constant $c = c(k,d) > 0$, $\kappa^{-1}([0,\varepsilon]^{(d-k)k})$ contains $B(H, c\,\varepsilon)$, where $H = \operatorname{span}\{e_1, \dots, e_k\}$ and $\{e_1, \dots, e_d\}$ is the canonical basis of $\mathbb{R}^d$. Indeed, let c be the constant provided by Lemma D.2 and consider $K \in B(H, c\,\varepsilon)$. Let $\kappa(K) = \{y^1, \dots, y^k\}$, so that $K = \operatorname{span}\{e_i + y^i : i = 1, \dots, k\}$. Applying Lemma D.2 with $u_i = e_i$ for $i = 1, \dots, k$, $v = e_j + y^j$ and $u_{k+1} = (v - Pv)/\|v - Pv\|_\infty$, where P denotes the orthogonal projection onto H, we get that $\angle(v, H) \geq c\ \|v - Pv\|_\infty$. With $\|v - Pv\|_\infty = \|y^j\|_\infty$ and the fact that $\angle(v, H) \leq \angle(K, H)$, we see that we have $\|y^j\|_\infty \leq \varepsilon$. This being true for all j, we have $\kappa(K) \in [0,\varepsilon]^{k(d-k)}$. We then apply the following result.

Lemma 6.3. *There is a constant $c = c(k,d) > 0$ such that, for all $\varepsilon > 0$ and for all $H \in \mathbb{G}(k,d)$, $\lambda(B(H, c\,\varepsilon)) \geq c\ \varepsilon^{(d-k)k}$.*

Lemma 6.3 is a direct consequence of Lemma B.2 in Appendix B and again the fact that $\lambda(\mathbb{G}(k,d)) = 1$.

Appendix

A Proof of Lemma 4.1

Lemma 4.1 is a conditional version of *The Coupon Collector's Problem* – see e.g. (11). We nevertheless provide here an elementary proof.

Let $K = N_n^{(r_0)}(\cup_{m\in\mathcal{M}} R_m)$. We know that $K \sim \mathrm{Bin}(n,p)$, where $p = |\mathcal{M}|\ \mu(R_{\mathbf{0}})$ with $|\mathcal{M}| \propto n^{\rho(r_0)}$ and

$$\mu(R_{\mathbf{0}}) = (\varepsilon')^k \cdot (\varepsilon/2)^{d-k} \cdot \prod_{\mathbf{s}\in\mathbf{S}\setminus\{\mathbf{0}\}} (\varepsilon_{|\mathbf{s}|}/(2\beta))^{d-k} \propto (\varepsilon')^k \varepsilon^{(d-k)w}.$$

This implies that $pn = c'|\mathcal{M}|$ for some $c' > 0$ not depending on n, by definition of ε and ε'. Let $c = c'/2$ and $c_0 \in (e^{-c}, 1)$, and also, to simplify notation, let $\ell = |\mathcal{M}|$ and $S = |\mathcal{M}| - |M|$. Because $|\mathcal{M}| \propto n^{\rho(r_0)}$, it is enough to show that $\mathbf{P}\{S > c_0\ \ell\} \to 0$ as $n \to \infty$.
We have

$$\mathbf{P}\{S > c_0\ \ell\} \le \mathbf{P}\{S > c_0\ \ell | K > c\ \ell\} + \mathbf{P}\{K > c\ \ell\},$$

with $\mathbf{P}\{K > c\ \ell\} \to 0$ as $n \to \infty$ by Lemma 3.3, and

$$\mathbf{P}\{S > c_0\ \ell | K > c\ \ell\} \le \mathbf{P}\{S > c_0\ \ell | K = \lceil c\ \ell\rceil\}.$$

Using Chebychev's inequality, we get

$$\mathbf{P}\{S > c_0\ \ell | K = \lceil c\ \ell\rceil\} \le \frac{\mathbf{var}\{S|K = \lceil c\ \ell\rceil\}}{(c_0\ \ell - \mathbf{E}\{S|K = \lceil c\ \ell\rceil\})^2}.$$

We know that for any non-negative integer k,

$$\mathbf{E}\{S|K = k\} = \ell(1 - 1/\ell)^k,$$

and

$$\mathbf{var}\{S|K = k\} = \ell((1 - 1/\ell)^k - (1 - 1/\ell)^{2k}) + \ell(\ell - 1)((1 - 2/\ell)^k - (1 - 1/\ell)^{2k}).$$

Therefore, when $\ell \to \infty$,

$$\mathbf{E}\{S|K = \lceil c\ \ell\rceil\} \backsim e^{-c}\ell,$$

and, for all ℓ,

$$\mathbf{var}\{S|K = \lceil c\ \ell\rceil\} \le c_1\ell,$$

so that, when ℓ is large,

$$\frac{\mathbf{var}\{S|K = \lceil c\ \ell\rceil\}}{(c_0\ \ell - \mathbf{E}\{S|K = \lceil c\ \ell\rceil\})^2} \le \frac{c_2}{\ell}.$$

Since ℓ is an increasing function of n that tends to infinity, we conclude that

$$\mathbf{P}\{S > c_0\ \ell\} \to 0, \quad n \to \infty.$$

□

B Coverings of $\mathbb{G}(k,d)$

Lemma B.1. *There is a constant $c = c(k,d) > 0$ such that, for all $\varepsilon > 0$, there is $H_1, \ldots, H_\ell \in \mathbb{G}(k,d)$ with $\ell > c\,\varepsilon^{-(d-k)k}$ and $B(H_i,\varepsilon) \cap B(H_j,\varepsilon) = \emptyset$ if $i \neq j$.*

Proof of Lemma B.1. Fix $\varepsilon > 0$ and consider

$$H_{\mathbf{n}} = \operatorname{span}\left\{ e_i + \varepsilon \sum_{j=k+1}^{d} n_{i,j} e_j : i = 1, \ldots, k \right\},$$

where $\mathbf{n} = (n_{i,j} : i = 1, \ldots, k; j = k+1, \ldots, d) \in ([0, 1/\varepsilon] \cap \mathbb{N})^{(d-k)k}$. As there are more than $1/2\ \varepsilon^{-(d-k)k}$ such $H_{\mathbf{n}}$'s, it suffices to prove that, for some constant $c > 0$,

$$\angle\,(H_{\mathbf{m}}, H_{\mathbf{n}}) \geq c\,\varepsilon,$$

as soon as $\mathbf{m} \neq \mathbf{n}$, for that would imply that the balls $B(H_{\mathbf{n}}, c/3\ \varepsilon)$ are disjoint when $\mathbf{n}$ runs through $([0, 1/\varepsilon] \cap \mathbb{N})^{(d-k)k}$.

Therefore, fix $\mathbf{m} \neq \mathbf{n}$, both in $([0, 1/\varepsilon] \cap \mathbb{N})^{(d-k)k}$. For $i = 1, \ldots, k$, let $u_i = e_i + \varepsilon \sum_{j=k+1}^{d} m_{i,j} e_j$ and $v_i = e_i + \varepsilon \sum_{j=k+1}^{d} n_{i,j} e_j$, where we assume, without loss of generality, that $u_1 \neq v_1$. Now, by definition

$$\angle\,(H_{\mathbf{m}}, H_{\mathbf{n}}) \geq \angle\,(v_1, \operatorname{span}\{u_1, \ldots, u_k\}).$$

To proceed further, we apply Lemma D.2 with $u_1, \ldots, u_k$, $u_{k+1} = (v_1 - u_1)/\|v_1 - u_1\|_\infty$ and $v = v_1$. It is straightforward to see that the conditions are satisfied, since in particular $v_1 = u_1 + \|v_1 - u_1\|_\infty u_{k+1}$. Hence, for a constant $c > 0$ depending only on k, d,

$$\angle\,(v_1, \operatorname{span}\{u_1, \ldots, u_k\}) \geq c\ \|v_1 - u_1\|_\infty.$$

To conclude, note that $\|v_1 - u_1\|_\infty \geq \varepsilon$. □

Lemma B.2. *There is a constant $c = c(k,d) > 0$ such that, for all $\varepsilon > 0$, there is $H_1, \ldots, H_\ell \in \mathbb{G}(k,d)$ with $\ell < c\,\varepsilon^{-(d-k)k}$ and $\mathbb{G}(k,d) \subset \bigcup_i B(H_i,\varepsilon)$.*

Proof of Lemma B.2. Let $e_1, \ldots, e_d$ be the canonical basis for $\mathbb{R}^d$ and let c_1 be the constant given by Lemma D.4. As in the proof of Lemma B.1, define

$$H_{\mathbf{n}}^{\sigma} = \operatorname{span}\left\{ e_{\sigma(i)} + \varepsilon \sum_{j=k+1}^{d} n_{i,j} e_{\sigma(j)} : i = 1, \ldots, k \right\},$$

where $\mathbf{n} = (n_{i,j} : i = 1, \ldots, k; j = k+1, \ldots, d) \in ([0, (c_1+1)/\varepsilon) \cap \mathbb{N})^{(d-k)k}$ and σ is a permutation of $\{1, \ldots, d\}$. There are no more than $c\,\varepsilon^{-(d-k)k}$ such subsets, where

$$c = \binom{d}{k}\ (c_1 + 1)^{(d-k)k}.$$

We now show that there is a constant $c = c(k,d) > 0$ such that, for all $H \subset \mathbb{G}(k,d)$, there exists such an $H_{\mathbf{n}}^{\sigma}$ satisfying $\angle\,(H_{\mathbf{n}}^{\sigma}, H) \leq c\,\varepsilon$.

So fix $H \subset \mathbb{G}(k,d)$. From Lemma D.4, it comes that there is σ, a permutation of $\{1,\dots,d\}$, and $\xi_{i,\sigma(j)} \in [0,c_1]$ for $i=1,\dots,k$ and $j=k+1,\dots,d$, such that

$$H = \operatorname{span}\left\{ e_{\sigma(i)} + \sum_{j=k+1}^{d} \xi_{i,\sigma(j)} e_{\sigma(j)} : i = 1,\dots,k \right\}.$$

For $i=1,\dots,k$ and $j=k+1,\dots,d$, define $n_{i,j}$ to be the entire part of $\xi_{i,\sigma(j)}/\varepsilon$, thus obtaining $\mathbf{n} = (n_{i,j}) \in ([0,(c_1+1)/\varepsilon) \cap \mathbb{N})^{(d-k)k}$. Applying Lemma D.1 with, for $i=1,\dots,k$,

$$u_i = e_{\sigma(i)} + \sum_{j=k+1}^{d} \xi_{i,\sigma(j)} e_{\sigma(j)},$$

and

$$v_i = e_{\sigma(i)} + \varepsilon \sum_{j=k+1}^{d} n_{i,j} e_{\sigma(j)},$$

we get $\angle(H^{\sigma}_{\mathbf{n}}, H) \leq c\,\varepsilon$ for some constant $c = c(k,d) > 0$. □

C Proof of Lemma 6.1

Only the last property defining $\mathcal{F}$ is non-trivial. Fix $f \in \mathcal{H}$, and let $f(x) = (x, g(x))$. Fix $s \in \{1,\dots,k\}$ and $x \in [0,1]^k$, and let $\mathbf{s}$ be the s^{th} canonical basis vector of $\mathbb{R}^k$. We have $\partial_s f(x) = (\mathbf{s}, \partial_s g(x))$.

Let $v = v_1 + v_2$ where $v_1 = (\mathbf{s}, 0)$ and $v_2 = (0, \partial_s g(x))$, and pick one vector $w \in \operatorname{span}\{\partial_t f(x) : t \neq s\}$.

We first show that

$$\langle v, w\rangle^2 \leq \frac{(d-k)\beta^2}{1+(d-k)\beta^2} \, \|v\|^2 \cdot \|w\|^2.$$

Indeed, since v_1 is orthogonal to w, we have $\langle v, w\rangle = \langle v_2, w\rangle$, so that, using the Cauchy-Schwartz inequality,

$$\langle v, w\rangle^2 \leq \|v_2\|^2 \cdot \|w\|^2.$$

We have $\|v\|^2 = \|v_1\|^2 + \|v_2\|^2$. Moreover, $\|v_1\|^2 = 1$ and $\|v_2\|^2 \leq (d-k)\beta^2$, since $\|\partial_s f(x)\|_\infty \leq \beta$. Conclude with

$$\begin{aligned} \|v_2\|^2 &= \left(1 - \frac{1}{\|v\|^2}\right) \|v\|^2 \\ &\leq \left(1 - \frac{1}{1+(d-k)\beta^2}\right) \|v\|^2. \end{aligned}$$

Since w is arbitrary in $\operatorname{span}\{\partial_t f(x) : t \neq s\}$, this shows that

$$\angle\left(\partial_s f(x), \operatorname{span}\{\partial_t f(x) : t \neq s\}\right) \geq \arccos\left(\sqrt{\frac{(d-k)\beta^2}{1+(d-k)\beta^2}}\right).$$

Furthermore,

$$\begin{aligned}\arccos\left(\sqrt{\frac{(d-k)\beta^2}{1+(d-k)\beta^2}}\right) &= \arctan\left(\frac{1}{\beta\sqrt{d-k}}\right)\\ &\geq \frac{1}{2\beta\sqrt{d-k}},\end{aligned}$$

where the last inequality comes from the fact that $\arctan(y) \geq y/2$ for $y \leq \pi/2$.

D Auxiliary Results in Euclidean Spaces

Lemma D.1. *Let c_1, c_2, c_3 be three positive constants. Let $u_1, \dots, u_k; v_1, \dots, v_k \in \mathbb{R}^d$ such that for all $i = 1, \dots, k$, $c_1 \leq \|v_i\|_\infty, \|u_i\|_\infty \leq c_2$, and, if $k \geq 2$,*

$$\angle\left(u_i, \operatorname{span}\{u_j : j \neq i\}\right) \geq c_3;$$

$$\angle\left(v_i, \operatorname{span}\{v_j : j \neq i\}\right) \geq c_3.$$

Then, for a constant c depending only on k, d, c_1, c_2, c_3,

$$\angle\left(\operatorname{span}\{u_i : i = 1, \dots, k\}, \operatorname{span}\{v_i : i = 1, \dots, k\}\right) \leq c \max_{i=1,\dots,k} \|v_i - u_i\|_\infty.$$

Proof of Lemma D.1. By multiplying the constants that appear in the Lemma by constants that depend only on d, we can assume that the conditions in the Lemma hold for the Euclidean norm. Throughout, let $\varepsilon = \max_{i=1,\dots,k} \|v_i - u_i\|_\infty$.

First assume that $u_1, \dots, u_k$ (resp. $v_1, \dots, v_k$) are orthonormal. Take

$$u = \sum_i \xi_i u_i \in \operatorname{span}\{u_i : i = 1, \dots, k\},$$

of norm equal to 1. Define

$$v = \sum_i \xi_i v_i \in \operatorname{span}\{v_i : i = 1, \dots, k\}.$$

We show that

$$\arccos(|\langle u, v\rangle|) = O(\varepsilon),$$

by showing that

$$\langle u, v\rangle = 1 + O(\varepsilon^2).$$

This comes from the fact that, since $\|u\| = \|v\| = 1$,

$$\langle u, v\rangle = 1 - \|u - v\|^2/2,$$

and

$$\|u - v\| \leq \sum_i |\xi_i| \, \|u_i - v_i\| \leq \sqrt{d}\, \varepsilon.$$

If $u_1, \dots, u_k$ (resp. $v_1, \dots, v_k$) are not orthonormal, we make them so. Define $a_1' = u_1$ and $a_1 = a_1'/\|a_1'\|$, and for $i = 2, \dots, k$, define

$$a_i' = u_i - \sum_{j=1}^{i-1} \langle u_i, a_j \rangle a_j,$$

and $a_i = a_i'/\|a_i'\|$. Similarly, define $b_1' = v_1$ and $b_1 = b_1'/\|b_1'\|$, and for $i = 2, \dots, k$, define

$$b_i' = v_i - \sum_{j=1}^{i-1} \langle v_i, b_j \rangle b_j,$$

and $b_i = b_i'/\|b_i'\|$.

We have, for $i = 1, \dots, k$, $c_1 \sin c_3 \leq \|a_i'\| \leq c_2$. Indeed, since a_i' is the difference between u_i and its orthogonal projection onto span $\{u_1, \dots, u_{i-1}\}$, it follows that

$$\|a_i'\| = \|u_i\| \ \sin \angle (u_i, \operatorname{span}\{u_1, \dots, u_{i-1}\}),$$

with

$$\angle (u_i, \operatorname{span}\{u_1, \dots, u_{i-1}\}) \geq \angle (u_i, \operatorname{span}\{u_j : j \neq i\}) \geq c_3 > 0.$$

In the same way, for $i = 1, \dots, k$, $c_1 \sin c_3 \leq \|b_i'\| \leq c_2$.

We also have $a_i' - b_i' = O(\varepsilon)$. We prove that recursively. First, $\|a_1' - b_1'\| = \|u_1 - v_1\| \leq \varepsilon$. Assume $a_{i-1}' - b_{i-1}' = O(\varepsilon)$. This implies $a_{i-1} - b_{i-1} = O(\varepsilon)$; indeed,

$$\begin{aligned} a_{i-1} - b_{i-1} &= \frac{\|b_{i-1}'\| a_{i-1}' - \|a_{i-1}'\| b_{i-1}'}{\|a_{i-1}'\| \|b_{i-1}'\|} \\ &\leq \frac{(\|a_{i-1}'\| + O(\varepsilon)) a_{i-1}' - \|a_{i-1}'\| (a_{i-1}' + O(\varepsilon))}{c^2} \\ &= O(\varepsilon). \end{aligned}$$

Now,

$$a_i' - b_i' = u_i - v_i - (\langle u_i, a_{i-1} \rangle a_{i-1} - \langle v_i, b_{i-1} \rangle b_{i-1}),$$

with $u_i - v_i = O(\varepsilon)$ and

$$\langle v_i, b_{i-1} \rangle b_{i-1} = \langle v_i, a_{i-1} + O(\varepsilon) \rangle (a_{i-1} + O(\varepsilon)) = \langle v_i, a_{i-1} \rangle a_{i-1} + O(\varepsilon).$$

So that

$$\langle u_i, a_{i-1} \rangle a_{i-1} - \langle v_i, b_{i-1} \rangle b_{i-1} = \langle u_i - v_i, a_{i-1} \rangle a_{i-1} + O(\varepsilon) = O(\varepsilon).$$

Hence, the recursion is satisfied.

We then apply the first part to $a_1, \dots, a_k$ and $b_1, \dots, b_k$. $\square$

Lemma D.2. *Fix c_1, c_2, c_3 three positive constants. Let $u_1, \dots, u_{k+1} \in \mathbb{R}^d$ such that for $i = 1, \dots, k+1$, $c_1 \leq \|u_i\|_\infty \leq c_2$, and*

$$\angle (u_i, \operatorname{span}\{u_j : j \neq i\}) \geq c_3.$$

Then, there is a positive constant c depending only on k, d, c_1, c_2, c_3 such that, for all $v = \sum_i \xi_i u_i$ with $\|v\|_\infty \leq c_2$,

$$\angle (v, \operatorname{span}\{u_i : i = 1, \dots, k\}) \geq c\ (|\xi_{k+1}| \wedge 1).$$

Proof of Lemma D.2. The proof is similar to that of Lemma D.1 above. Again, we may work with the Euclidean norm instead of the supnorm.

First assume that $u_1, \ldots, u_{k+1}$ are orthonormal. Take $v = \sum_i \xi_i u_i$ of norm equal to $\sqrt{\sum_i \xi_i^2} \leq c_2$. Let $Pv = \sum_{i=1}^k \xi_i u_i$, the orthonormal projection of v onto $\mathrm{span}\,\{u_1, \ldots, u_k\}$. By definition,

$$\angle\,(v, \mathrm{span}\,\{u_1, \ldots, u_k\}) = \arccos\left(\frac{\langle v, Pv\rangle}{\|v\|\|Pv\|}\right).$$

We then conclude with

$$\frac{\langle v, Pv\rangle}{\|v\|\|Pv\|} = \sqrt{1 - \frac{\xi_{k+1}^2}{\|v\|^2}} = 1 + O(\xi_{k+1}^2).$$

In general, we first make $u_1, \ldots, u_{k+1}$ orthonormal as we did in the proof of Lemma D.1, except in reverse order, meaning that $a_{k+1} = u_{k+1}/\|u_{k+1}\|$. Since for all $v = \sum_i \xi_i u_i = \sum_i \gamma_i a_i$, $|\gamma_{k+1}| = |\xi_{k+1}|\|u_{k+1}\| \geq c_1\ |\xi_{k+1}|$, we can apply the first part to $a_1, \ldots, a_{k+1}$. □

Lemma D.3. *Let $e_1, \ldots, e_d$ be the canonical basis of $\mathbb{R}^d$. There is a constant $c = c(k, d) > 0$ such that, for every $u_1, \ldots, u_k$, orthonormal set of vectors in $\mathbb{R}^d$, there exists a permutation σ of $\{1, \ldots, d\}$ such that*

$$|\langle u_i, e_{\sigma(i)}\rangle| \geq c, \quad \forall i = 1, \ldots, k.$$

Proof of Lemma D.3. We prove Lemma D.3 by recursion on k. For $k = 1$, we may choose $c(1, d) = 1/\sqrt{d}$, since u_1 is of norm 1. Suppose the result is true at $k - 1$, and consider the case at k. Without loss of generality, we may assume that

$$|\langle u_i, e_i\rangle| \geq c(k-1, d), \quad \forall i = 1, \ldots, k-1.$$

We need to show that there is a constant $c_1 > 0$ and $j \in \{k, \ldots, d\}$ such that

$$|\langle u_k, e_j\rangle| \geq c_1.$$

As Lemma D.3 implies Lemma D.4, we can use the latter at $k - 1$ to get vectors $e_i + v_i$, $i = 1, \ldots, k-1$, with $v_i \in \mathrm{span}\,\{e_k, \ldots, e_d\}$, and $\|v_i\|_\infty \leq c_2$, such that

$$\mathrm{span}\,\{e_1 + v_1, \ldots, e_{k-1} + v_{k-1}\} = \mathrm{span}\,\{u_1, \ldots, u_{k-1}\}.$$

Let $\xi = \max_{j=k,\ldots,d} |\langle u_k, e_j\rangle|$. For all $i = 1, \ldots, k-1$, we have

$$|\langle u_k, v_i\rangle| \leq (d-k)c_2\xi.$$

Now, since for all $i = 1, \ldots, k-1$, $\langle u_k, e_i + v_i\rangle = 0$, we also have

$$|\langle u_k, e_i\rangle| \leq (d-k)c_2\xi.$$

Since there is $i = 1, \ldots, d$ such that $|\langle u_k, e_i\rangle| \geq 1/\sqrt{d}$, we must have

$$\xi \geq \frac{1}{\sqrt{d}} \vee \frac{1}{\sqrt{d}(d-k)c_2}.$$

Conclude by calling the right handside c_3 and letting

$$c(k, d) = c(k-1, d) \wedge c_3.$$

□

Lemma D.4. *Let $e_1, \ldots, e_d$ be the canonical basis of $\mathbb{R}^d$. There is a constant $c = c(k, d) > 0$ such that, for $u_1, \ldots, u_k$ any orthonormal set of vectors in $\mathbb{R}^d$, there exists a permutation σ of $\{1, \ldots, d\}$, such that*

$$\operatorname{span}\{u_1, \ldots, u_k\} = \operatorname{span}\left\{e_{\sigma(1)} + v_1, \ldots, e_{\sigma(k)} + v_k\right\},$$

where, for all $i = 1, \ldots, k$,

$$v_i \in \operatorname{span}\{e_\sigma(j) : j = k+1, \ldots, d\},$$

and $\|v_i\|_\infty \leq c$.

Proof of Lemma D.4. Applying Lemma D.3, there is $c_1 > 0$ and a permutation σ such that

$$|\langle u_i, e_{\sigma(i)}\rangle| \geq c_1, \quad \forall i = 1, \ldots, k.$$

Without loss of generality, suppose $\sigma = \mathrm{id}$.

We now triangulate the matrix with column vectors $u_1, \ldots, u_k$. In other words, we consider $\{u_i' : i = 1, \ldots, k\}$, where u_i' is the orthogonal projection of u_i onto $\operatorname{span}\{e_i, e_{k+1}, \ldots, e_d\}$. For all $i = 1, \ldots, k$, we have $u_i' = \xi_i e_i + w_i$, where $|\xi_i| \geq c_1$ and $w_i \in \operatorname{span}\{e_{k+1}, \ldots, e_d\}$ with $\|w_i\|_\infty \leq 1$. Define $v_i = w_i/\xi_i$ and conclude with the fact that

$$\operatorname{span}\left\{u_1', \ldots, u_k'\right\} = \operatorname{span}\{u_1, \ldots, u_k\}.$$

□

References

[1] P.-A. Absil, A. Edelman, and P. Koev. On the largest principal angle between random subspaces. *Linear Algebra Appl.*, 414 (2006), no. 1, 288–294. MR2209246

[2] E. Arias-Castro, D. L. Donoho, X. Huo, and C. Tovey. Connect-the-dots: How many random points can a regular curve pass through? *Adv. in Appl. Probab.*, 37 (2005), no. 3, 571–603. MR2156550

[3] E. Arkin, J. Mitchell, and G. Narasimhan. Resource-constrained geometric network optimization. In *Proc. of ACM Symposium on Computational Geometry, Minneapolis*, (1997), 307–316.

[4] B. Awerbuch, Y. Azar, A. Blum, and S. Vempala. New approximation guarantees for minimum-weight k-trees and prize-collecting salesmen. *SIAM J. Comput.*, 28 (1999), no. 1, 254–262. MR1630453

[5] B. DasGupta, J. Hespanha, and E. Sontag. Computational complexities of honey-pot searching with local sensory information. In *2004 American Control Conference (ACC 2004)*, pages 2134–2138. 2004.

[6] D. Field, A. Hayes, and R. Hess. Contour integration by the human visual system: evidence for a local association field. *Vision Research*, 33(2):173–193, 1993.

[7] G. H. Golub and C. F. Van Loan. *Matrix computations.* Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996. MR1417720

[8] X. Huo, D. Donoho, C. Tovey, and E. Arias-Castro. Dynamic programming methods for 'connecting the dots' in scattered point clouds. *INFORMS J. Comput.*, 2005. To appear.

[9] A. N. Kolmogorov. *Selected works of A. N. Kolmogorov. Vol. III*, volume 27 of *Mathematics and its Applications (Soviet Series).* Kluwer Academic Publishers Group, Dordrecht, 1993. MR1228446

[10] V. D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov. MR856576

[11] F. Mosteller. *Fifty challenging problems in probability with solutions.* Addison-Wesley Publishing Co., Inc., Reading, Mass.-London, 1965. MR397810

[12] G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics.* John Wiley & Sons Inc., New York, 1986. MR838963