On the Additive Properties of the Fat-Shattering Dimension

Ohad Asor, Hubert Haoyang Duan, Aryeh Kontorovich

***Abstract*—The properties of the VC-dimension under various compositions are well-understood, but this is much less the case for classes of continuous functions. In this paper, we show that a commonly used scale-sensitive dimension, V_γ, is much less well-behaved under Minkowski summation than its VC cousin, while the fat-shattering dimension retains some compositional similarity to the VC-dimension. As an application, we analyze the fat-shattering dimension of trigonometric functions and series.**

I. Introduction

Combinatorial dimensions play a central role in learning theory, as they allow one to precisely characterize the learnable function classes [1], [2], [15], [20]. Unfortunately, in all but the simplest cases, these dimensions are quite difficult to compute exactly. Hence, it is useful to be able to estimate the combinatorial dimension of a complicated function family in terms of its simpler constituents.

In the case of binary-valued function classes, the relevant combinatorial parameter is the VC-dimension, which is fairly well-behaved under addition. In particular, suppose that $\mathcal{F}, \mathcal{G} : \Omega \to \{-1, 1\}$ are two concept classes and define $\mathcal{H} : \Omega \to \{-1, 1\}$ by

$$\mathcal{H} = \{\operatorname{sgn}(f(x) + g(x)) : f \in \mathcal{F}, g \in \mathcal{G}\}.$$

Then [4], [12]

$$\operatorname{VCdim}(\mathcal{H}) = O(\operatorname{VCdim}(\mathcal{F}) + \operatorname{VCdim}(\mathcal{G})). \quad (1)$$

Some lower bounds on the VC-dimension of compositions are also known [10], [11]. More generally, for $k \geq 2$, given any concept classes $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_k$ and function $u : \{-1, 1\}^k \to \{-1, 1\}$, one can define the concept class $u(\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_k)$ as the set consisting of all functions $u(f_1, f_2, \ldots, f_k) : \Omega \to \{-1, 1\}$ for

$$u(f_1, f_2, \ldots, f_k)(x) = u(f_1(x), f_2(x), \ldots, f_k(x)), \quad (2)$$

where $f_i \in \mathcal{F}_i$, $i = 1, 2, \ldots, k$. Vidyasagar in [21] provides an upper bound on the VC-dimension of this composition class in terms of the dimensions of the individual concept classes.

Theorem 1 ([21])**.** *If* $\operatorname{VCdim}(\mathcal{F}_i) < \infty$ *for all* $i = 1, 2, \ldots, k$*, then there is a finite* $\alpha = \alpha(k)$*, depending only on* k*, such that*

$$\operatorname{VCdim}(u(\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_k)) < \alpha \max_{i=1}^{k} \operatorname{VCdim}(\mathcal{F}_i).$$

Real-valued functions admit various analogues of the VC-dimension, going by the general name of *scale-sensitive dimensions*. Intuitively, these encorporate a measure of "margin" or "scale" into their notion of shattering (a formal definition will be given below). Unlike VC-dimension, compositional properties for various scale-sensitive dimensions are far less well understood. In this paper, we explore this gap and study the additive properties of generalized versions of the VC-dimension for function classes. In Section IV, we show that some dimensions do not possess a property analogous to (1), while others are far more well-behaved under Minkowski summation — and even satisfy a generalized version of Theorem 1 [8]. Finally, in Section V we analyze the fat-shattering dimension of trigonometric functions and series.

II. Related work

Since the seminal work of [1], [2], it has been known that, modulo measure-theoretic technicalities, a class of continuous-valued functions admits distribution-free risk bounds if and only if its scale-sensitive dimension is finite at each scale. This generalizes the analogous characterization of PAC-learnable Boolean function classes as those having a finite VC-dimension [5]. Regarding the latter, one commonly encounters situations where a complicated function class is constructed by combining simpler hypotheses via basic operations; examples include neural networks [4] and boosting [12]. This was the primary motivation behind results such as (1), as well as those in this paper. A precursor to the scale-sensitive dimensions was Pollard's *pseudo-dimension* $\operatorname{Pdim}(\cdot)$ [18], which, while capable of providing simple upper bounds [17], is too crude to characterize distribution-free convergence. One way of computing $\operatorname{Pdim}(\mathcal{F})$ is to calculate the VC-dimension of the set of all mappings $(x,t) \mapsto \mathbb{1}_{\{f(x)>t\}}$, indexed by $f \in \mathcal{F}$ [20]. This characterization implies that $\mathrm{Pdim}(\mathcal{F}+\mathcal{G}) \le O(\mathrm{Pdim}(\mathcal{F}) + \mathrm{Pdim}(\mathcal{G}))$.

Other approaches to learning continuous-valued functions by examining their discretized behavior include [22], [23]. A powerful alternative framework for analyzing the complexity of continuous function classes is that of *Rademacher complexity* [3], [16]. We will not define here the Rademacher complexity $R(\cdot)$ of a function class $\mathcal{F}$, but note in passing that it is rather well-behaved with respect to Minkowski addition and scalar multiplication [6]:

$$R(\alpha\mathcal{F}+\beta\mathcal{G}) \le |\alpha|R(\mathcal{F}) + |\beta|R(\mathcal{G}).$$

A recent example where Rademacher complexity is used to control approximation error may be found in [13].

III. Definitions

Let $\mathcal{F}$ be a collection of functions $f : \Omega \to \mathbb{R}$ and recall the definition of the V_γ and fat-shattering dimensions [1], [2]: a set $S \subset \Omega$ is said to be V_γ-*shattered by* $\mathcal{F}$ if there exists some constant $r \in \mathbb{R}$ such that for each label assignment $y \in \{-1,1\}^S$ there is an $f \in \mathcal{F}$ satisfying

$$y(x)(f(x)-r) \ge \gamma > 0 \tag{3}$$

for all $x \in S$. The set S is γ-*fat shattered by* $\mathcal{F}$, on the other hand, if there exists some *function* $r : S \to \mathbb{R}$, called the *witness of shattering*, such that for each label assignment $y \in \{-1,1\}^S$ there is an $f \in \mathcal{F}$ satisfying

$$y(x)(f(x)-r(x)) \ge \gamma > 0 \tag{4}$$

for all $x \in S$. The V_γ *dimension*, denoted by $V_\gamma(\mathcal{F})$, and the γ-*fat shattering dimension*, denoted by $\mathrm{fat}_\gamma(\mathcal{F})$, *of* $\mathcal{F}$ are respectively the cardinality of the largest set V_γ-shattered and γ-fat shattered by $\mathcal{F}$. If there exist sets of arbitrarily large sizes that are V_γ-shattered, or γ-fat shattered, by $\mathcal{F}$, then $V_\gamma(\mathcal{F}) = \infty$, or $\mathrm{fat}_\gamma(\mathcal{F}) = \infty$, respectively. The only difference between the V_γ and the fat-shattering dimensions is that r is required to be a constant for V_γ-shattering while it can be a real-valued function on S for γ-fat shattering. As a result,

$$V_\gamma(\mathcal{F}) \le \mathrm{fat}_\gamma(\mathcal{F}), \tag{5}$$

for any function class $\mathcal{F}$.

Recall from basic analysis that a function $u : X \to Y$ between two metric spaces (X, d_X) and (Y, d_Y) is *uniformly continuous* if for every $\epsilon > 0$, there exists $\delta = \delta(\epsilon)$ such that for all $x, x' \in X$,

$$d_X(x,x') < \delta \Rightarrow d_Y(u(x), u(x')) < \epsilon.$$

If $\mathcal{F}, \mathcal{G}$ are two sets of real-valued functions defined on the same domain, $\mathcal{F}+\mathcal{G}$ denotes their *Minkowski sum*:

$$\mathcal{F}+\mathcal{G} = \{h = f + g : f \in \mathcal{F}, g \in \mathcal{G}\}.$$

The natural numbers and integers are denoted by $\mathbb{N}$ and $\mathbb{Z}$, respectively. For $\boldsymbol{c} = (c_1, c_2, \ldots) \in \mathbb{R}^{\mathbb{N}}$, we write $\|\boldsymbol{c}\|_1$ to denote $\sum_{i\in\mathbb{N}} |c_i|$. Standard asymptotic order-of-magnitude notation $O(\cdot)$ and $\Omega(\cdot)$ is used.

IV. Additive properties

The next result shows that there is no V_γ-shattering analogue of (1) by exhibiting function families $\mathcal{F}, \mathcal{G}$ such that $V_\gamma(\mathcal{F}), V_\gamma(\mathcal{G})$ are small while $V_\gamma(\mathcal{F}+\mathcal{G})$ is large:

Theorem 2. *There exist function classes $\mathcal{F}, \mathcal{G}$ such that for all $\gamma > 0$, $V_\gamma(\mathcal{F}) = V_\gamma(\mathcal{G}) = 1$ and $V_\gamma(\mathcal{F}+\mathcal{G}) = \infty$.*

Proof: For $-\infty < a < b < \infty$, let $\mathcal{F}$ be the collection of all increasing functions on $[a,b]$ and let $\mathcal{G}$ be the collection of all decreasing functions on $[a,b]$. Clearly, $V_\gamma(\mathcal{F}) = V_\gamma(\mathcal{G}) = 1$ for all $\gamma > 0$. However, $\mathcal{H} = \mathcal{F}+\mathcal{G} = BV[a,b]$ is the collection of functions of bounded variation on $[a,b]$ and is easily seen to have $V_\gamma(\mathcal{H}) = \infty$ for all $\gamma > 0$, see e.g. [19]. ■

On the other hand, the fat-shattering dimension is significantly better behaved under Minkowski summation. More generally, it admits a version of Theorem 1:

Theorem 3 **([8]).** *Let $\gamma > 0$, $k \ge 2$, $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_k$ be function classes consisting of functions $f : \Omega \to [-1,1]$, and $u : [-1,1]^k \to [-1,1]$ be uniformly continuous. Then there are $0 < \alpha, \beta < \infty$, depending only on u, k and γ, such that*

$$\mathrm{fat}_\gamma(u(\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_k)) \le \alpha \sum_{i=1}^{k} \mathrm{fat}_\beta(\mathcal{F}_i),$$

where $u(\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_k)$ consists of all functions $u(f_1, f_2, \ldots, f_k) : \Omega \to [-1,1]$ as defined in (2).

Since addition is a uniformly continuous operation, we have the following result:

Corollary 4. *Let $\mathcal{F}$ and $\mathcal{G}$ consist of functions $f : \Omega \to [-0.5, 0.5]$ and $\mathcal{H} = \mathcal{F}+\mathcal{G}$ be their Minkowski* *sum. Then there is a constant* $0 < \alpha < \infty$ *and a function* $\beta : (0, \infty) \to (0, \infty)$ *such that*

$$\mathrm{fat}_\gamma(\mathcal{H}) \le \alpha(\mathrm{fat}_{\beta(\gamma)}(\mathcal{F}) + \mathrm{fat}_{\beta(\gamma)}(\mathcal{G})).$$

It is instructive to contrast Corollary 4 with Theorem 2. Moreover, the class of functions on $[a, b]$ with bounded total variation norm has a well-behaved fat-shattering dimension, and thus the V_γ dimension as well, by (5). Recall the definition of *total variation* of a function $f : [a, b] \to \mathbb{R}$:

$$V_a^b(f) = \sup_P \sum_{i=0}^{|P|-1} |f(x_{i+1}) - f(x_i)|, \tag{6}$$

where the supremum is over the set of all finite partitions $P = (a = x_0, x_1, \ldots, x_{|P|} = b)$ of $[a, b]$.

Theorem ([19])**.** *Let* $\mathcal{F}$ *be the collection of all* $f : [a, b] \to \mathbb{R}$ *such that* $V_a^b(f) \le L$*. Then*

$$V_\gamma(\mathcal{F}) = \mathrm{fat}_\gamma(\mathcal{F}) = 1 + \left\lfloor \frac{L}{2\gamma} \right\rfloor.$$

In contradistinction to Theorem 2, it may be the case that $V_\gamma(\mathcal{F})$ and $V_\gamma(\mathcal{G})$ are large while $V_\gamma(\mathcal{F} + \mathcal{G})$ is small. We state the following simple lemma without proof:

Lemma 5. *Let* $\mathcal{F}$ *be a collection of differentiable functions* $f : \mathbb{R} \to \mathbb{R}$ *such that for each* $f \in \mathcal{F}$*, the derivative* f' *vanishes on at most* L *points. Then* $V_\gamma(\mathcal{F}) \le L + 2$ *for all* $\gamma > 0$*.*

Theorem 6. *There exist function classes* $\mathcal{F}, \mathcal{G}$ *such that for all* $\gamma > 0$*,* $V_\gamma(\mathcal{F}) = V_\gamma(\mathcal{G}) = \infty$ *and* $V_\gamma(\mathcal{F} + \mathcal{G}) \le 3$*.*

Proof: Let $\mathcal{F}$ be the collection of all differentiable $f : \mathbb{R} \to \mathbb{R}$ such that $f' < -1$ on $(-\infty, 0)$ and $-1 \le f' \le 1$ on $(0, \infty)$. Let $\mathcal{G}$ be the collection of all differentiable $g : \mathbb{R} \to \mathbb{R}$ such that $-1 \le g' \le 1$ on $(-\infty, 0)$ and $g' > 1$ on $(0, \infty)$. Consider an $h \in \mathcal{F} + \mathcal{G}$. By construction, $h' < 0$ on $(-\infty, 0)$ and $h' > 0$ on $(0, \infty)$. Since h' only vanishes at $x = 0$, the claim follows by Lemma 5. ■

Remark 7. An analogous construction exists for the VC-dimension. Let $\mathcal{F} : \Omega_1 \to \{-1, 1\}$ and $\mathcal{G} : \Omega_2 \to \{-1, 1\}$ be two concept classes defined over two disjoint sets Ω_1, Ω_2. Put $\Omega = \Omega_1 \cup \Omega_2$ and define $\bar{\mathcal{F}} : \Omega \to \{-1, 1\}$ by

$$\bar{f}(x) = \begin{cases} f(x), & x \in \Omega_1 \\ 1, & x \in \Omega_2 \end{cases},$$

and $\bar{\mathcal{G}} : \Omega \to \{-1, 1\}$ analogously. Define

$$\mathcal{H} = \bar{\mathcal{F}} \vee \bar{\mathcal{G}} = \left\{ h = \bar{f} \vee \bar{g} : \bar{f} \in \bar{\mathcal{F}}, \bar{g} \in \bar{\mathcal{G}} \right\}.$$

Then $\mathrm{VCdim}(\mathcal{H}) = 0$, regardless of how large $\mathrm{VCdim}(\bar{\mathcal{F}})$ or $\mathrm{VCdim}(\bar{\mathcal{G}})$ might be.

However, such construction of having infinite dimension for $\mathcal{F}$ and $\mathcal{G}$ but finite dimension for $\mathcal{F} + \mathcal{G}$ cannot exist for the γ-fat shattering dimension.

Theorem 8. *For any* $\gamma > 0$*, if* $\mathcal{F}$ *and* $\mathcal{G}$ *are non-empty function classes and* $\mathrm{fat}_\gamma(\mathcal{F}) = \infty$ *or* $\mathrm{fat}_\gamma(\mathcal{G}) = \infty$*, then* $\mathrm{fat}_\gamma(\mathcal{F} + \mathcal{G}) = \infty$*.*

Proof: Without loss of generality, let $\mathrm{fat}_\gamma(\mathcal{F}) = \infty$ and $n \in \mathbb{N}$. There exists some set $S \subseteq \Omega$, with $|S| = n$, which is γ-fat shattered by $\mathcal{F}$ with witness of shattering $r : S \to \mathbb{R}$. Then, $\mathcal{F} + \mathcal{G}$ also γ-shatters S: pick any $g \in \mathcal{G}$ and consider the witness of shattering $r + g : S \to \mathbb{R}$; given a label assignment $y \in \{-1, 1\}^S$, there exists $f \in \mathcal{F}$ such that

$$y(x)(f(x) - r(x)) \ge \gamma > 0,$$

so for $f + g \in \mathcal{F} + \mathcal{G}$,

$$y(x)([f(x) + g(x)] - [r(x) + g(x)]) \ge \gamma > 0.$$

Hence, $\mathrm{fat}_\gamma(\mathcal{F} + \mathcal{G}) = \infty$. ■

V. Trigonometric functions and series

In this section, we give explicit upper and lower estimates on the fat-shattering dimension of trigonometric functions and series. Let $\mathcal{F}_{\sin}$ be the family of functions $f : \mathbb{R} \to \mathbb{R}$ defined by

$$\mathcal{F}_{\sin} = \{f(x) = \sin(\alpha x) : \alpha \in \mathbb{R}\}$$

and define the concept class $\mathrm{sgn}(\mathcal{F}_{\sin}) : \mathbb{R} \to \{-1, 1\}$ by

$$\mathrm{sgn}(\mathcal{F}_{\sin}) = \{\mathrm{sgn}(f) : f \in \mathcal{F}\}.$$

A well-known construction due to E. Levin and J. S. Denker [7], [20] shows that $\mathrm{VCdim}(\mathrm{sgn}(\mathcal{F}_{\sin})) = \infty$, providing an example of a concept class parametrized by a single real number with infinite VC-dimension.

We extend this construction to the fat-shattering dimension:

Theorem 9. *For all* $0 < \gamma < 1$*,*

$$\mathrm{fat}_\gamma(\mathcal{F}_{\sin}) = \infty.$$

Proof: Our proof is analogous to the construction in [7, Section 2.3]. Let $n \in \mathbb{N}$ and $0 < \gamma < 1$ be given. Put $\beta = (\sin^{-1} \gamma)/\pi$ and

$$b = 2\left(2\left\lceil \frac{1}{1/2 - \beta} \right\rceil + 1\right).$$

Define $\boldsymbol{x} \in \mathbb{R}^n$ by

$$x_i = b^{-i}, \qquad i = 1, \dots, n. \tag{7}$$

For each $\boldsymbol{y} \in \{-1, 1\}^n$, define

$$\alpha = \frac{b}{2} + \frac{1}{2} \sum_{i=1}^{n} (b - y_i - 1) b^i. \tag{8}$$

It is straightforward to verify that

$$y_i \sin(\pi \alpha x_i) \geq \gamma, \qquad i = 1, \dots, n.$$

■

Obviously, $\operatorname{fat}_\gamma(\mathcal{F}_{\sin}) = 0$ for $\gamma > 1$. The case of $\gamma = 1$ requires a separate treatment:

Theorem 10.

$$\operatorname{fat}_1(\mathcal{F}_{\sin}) = 1.$$

Proof: Clearly, $\mathcal{F}_{\sin}$ can shatter a single point at $\gamma = 1$. Suppose, contrary to our claim, that $\mathcal{F}_{\sin}$ can shatter some two-point set, $\{x_1, x_2\}$ at $\gamma = 1$. This implies the existence of $\alpha_1, \alpha_2 \in \mathbb{R}$ such that

$$\begin{aligned}
\sin(\alpha_1 x_1) = +1 \quad &\Longleftrightarrow \quad \alpha_1 x_1 \in (2\mathbb{Z} + 1/2)\pi \\
\sin(\alpha_1 x_2) = +1 \quad &\Longleftrightarrow \quad \alpha_1 x_2 \in (2\mathbb{Z} + 1/2)\pi \\
\sin(\alpha_2 x_1) = +1 \quad &\Longleftrightarrow \quad \alpha_2 x_1 \in (2\mathbb{Z} + 1/2)\pi \\
\sin(\alpha_2 x_2) = -1 \quad &\Longleftrightarrow \quad \alpha_2 x_2 \in (2\mathbb{Z} + 3/2)\pi.
\end{aligned}$$

Rearranging, we have

$$\frac{x_1}{x_2} \in \frac{2\mathbb{Z} + 1/2}{2\mathbb{Z} + 1/2} \cap \frac{2\mathbb{Z} + 1/2}{2\mathbb{Z} + 3/2}. \tag{9}$$

In particular, (9) implies the existence of $k, \ell, m, n \in \mathbb{Z}$ such that

$$\frac{4k + 1}{4\ell + 1} = \frac{4m + 1}{4n + 3},$$

which is impossible. ■

A partial converse to Theorem 9 is that restricting both the range of x and the frequency α yields finite fat-shattering dimension:

Theorem 11. *Let $\mathcal{F}$ be the family of functions $f : [0, M] \to \mathbb{R}$ defined by*

$$\mathcal{F} = \{f(x) = \sin(\alpha x) : \alpha \in [-A, A]\}$$

for some $0 < M, A < \infty$. Then

$$\operatorname{fat}_\gamma(\mathcal{F}) \leq \left\lfloor \frac{MA}{2\gamma} \right\rfloor + 1, \qquad 0 < \gamma < 1. \tag{10}$$

Proof: Since $|f'| \leq A$ for all $f \in \mathcal{F}$, it suffices to prove that (10) holds when $\mathcal{F}$ is the collection of all A-Lipschitz functions on $[0, M]$. Appealing to [14, Theorem 2], we have

$$\operatorname{fat}_\gamma(\mathcal{F}) \leq \mathscr{M}(2\gamma/A),$$

where $\mathscr{M}(\cdot)$ is the packing number of $[0, M]$. Since $\mathscr{M}(\varepsilon) = \lfloor M/\varepsilon \rfloor + 1$, the claim follows. ■

Remark 12. If either the range of x or the frequency α is unbounded, we have $\operatorname{fat}_\gamma(\mathcal{F}) = \infty$ for $0 < \gamma < 1$. This is easily seen from the construction in Theorem 9. Suppose the range of x is $[0, 1]$ and α is unbounded. Instead of the $\boldsymbol{x}$ and α constructed in (7) and (8), respectively, put $\boldsymbol{x}' = \boldsymbol{x}/b$ and $\alpha' = b\alpha$. The case of bounded α and unbounded x is handled analogously.

Our observations regarding sine functions can be extended to trigonometric series.

Theorem 13. *Let $I \subseteq \mathbb{R}$ be a (possibly unbounded) segment and let $A, L \in (0, \infty]$. Let $\mathcal{F}_{I,A,L}$ be the family of functions $f : I \to \mathbb{R}$ defined by*

$$\begin{aligned}
\mathcal{F}_{I,A,L} \quad = \quad &\Big\{f(x) = \sum_{i=1}^{\infty} c_i \sin(\alpha_i x) : \boldsymbol{c} \in \mathbb{R}^{\mathbb{N}}, \\
&\|\boldsymbol{c}\|_1 \leq L,\ \boldsymbol{\alpha} \in [-A, A]^{\mathbb{N}}\Big\}.
\end{aligned}$$

Then

$$\operatorname{fat}_\gamma(\mathcal{F}_{I,A,L}) \begin{cases} = 0, & \gamma > L \\ = 1, & \gamma = L \\ \leq \left\lfloor \frac{|I| L A}{2\gamma} \right\rfloor + 1, & |I| < \infty, \\ & \gamma < L < \infty, A < \infty \\ = \infty, & \gamma < L, \infty \in \{|I|, A\}, \end{cases} \tag{11}$$

where $|I|$ is the length (Lebesgue measure) of I.

Proof: The first relation is obvious, since $\left|\sum_{i \in \mathbb{N}} c_i \sin(\alpha_i x)\right| \leq \|\boldsymbol{c}\|_1 \leq L$. To prove the second relation, note that $\left|\sum_{i \in \mathbb{N}} c_i \sin(\alpha_i x)\right| = L$ if and only if $|\sin(\alpha_i x)| = 1$ for each $i \in \mathbb{N}$. Hence, the argument from Theorem 10 can be applied termwise to show that $\mathcal{F}_{I,A,L}$ cannot shatter two points at $\gamma = L$. To prove the third relation, we observe that $\|\boldsymbol{c}\|_1 < \infty$ implies that $\sum c_i \sin(\alpha_i x)$ converges pointwise. Furthermore, there is no loss of generality in assuming that each $c_i > 0$ (since the corresponding α_i can be replaced by its negative). Hence, $\|\boldsymbol{c}\|_1^{-1} \sum_{i \in \mathbb{N}} c_i \sin(\alpha_i x)$ is a convex combination of A-Lipschitz functions defined on a compact domain and therefore is itself Lipschitz with constant at most A. Thus, the packing bound for Lipschitz functions invoked in the proof of Theorem 11 applies here as well. The fourth relation is an immediate consequence of Remark 12, since in particular $\mathcal{F}_{I,A,L}$ contains functions of the form $f(x) = L \sin(\alpha x)$. ■

Remark 14. The analysis above can be extended to more general trigonometric series of the form $\sum a_i \sin(\alpha_i x) + \sum b_i \cos(\beta_i x)$ in a straightforward fashion.

VI. Conclusion

We have investigated the behavior of various scale-sensitive dimensions under Minkowski addition, providing some positive and negative results. An intriguing future research direction would be to make the dependence of α, β on u, k, and γ in Theorem 3 more explicit. This could be of relevance, for example, to boosted regression [9] or continuous-output neural networks.

Acknowledgements

We thank Noga Alon for helpful correspondence and discussions.

References

[1] Noga Alon, Shai Ben-David, Nicolò Cesa-Bianchi, and David Haussler. Scale-sensitive dimensions, uniform convergence, and learnability. *Journal of the ACM*, 44(4):615–631, 1997.

[2] Peter L. Bartlett and Philip M. Long. Prediction, learning, uniform convergence, and scale-sensitive dimensions. *J. Comput. Syst. Sci.*, 56(2):174–190, 1998.

[3] Peter L. Bartlett and Shahar Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3:463–482, 2002.

[4] Eric B. Baum and David Haussler. What size net gives valid generalization? *Neural Comput.*, 1(1):151–160, 1989.

[5] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *J. Assoc. Comput. Mach.*, 36(4):929–965, 1989.

[6] Stéphane Boucheron, Olivier Bousquet, and Gábor Lugosi. Theory of classification: a survey of some recent advances. *ESAIM: Probability and Statistics*, 9:323–375, 2005.

[7] Christopher J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2(2):121–167, 1998.

[8] Hubert Haoyang Duan. Bounding the fat shattering dimension of a composition function class built using a continuous logic connective. *Waterloo Mathematics Review*, 2(1):4–20, 2012.

[9] Nigel Duffy and David Helmbold. Boosting methods for regression. *Machine Learning*, 47:153–200, 2002.

[10] David Eisenstat. k-fold unions of low-dimensional concept classes. *Inf. Process. Lett.*, 109(23-24):1232–1234, 2009.

[11] David Eisenstat and Dana Angluin. The VC dimension of k-fold union. *Inf. Process. Lett.*, 101(5):181–184, 2007.

[12] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55(1):119–139, 1997.

[13] Giorgio Gnecco and Marcello Sanguineti. Approximation error bounds via Rademacher's complexity. *Appl. Math. Sci. (Ruse)*, 2(1-4):153–176, 2008.

[14] Lee-Ad Gottlieb, Leonid Kontorovich, and Robert Krauthgamer. Efficient classification for metric data. In *COLT*, pages 433–440, 2010.

[15] Micheal Kearns and Umesh Vazirani. *An Introduction to Computational Learning Theory*. The MIT Press, 1997.

[16] Vladimir Koltchinskii and Dmitriy Panchenko. Empirical margin distributions and bounding the generalization error of combined classifiers. *Ann. Statist.*, 30(1):1–50, 2002.

[17] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations Of Machine Learning*. The MIT Press, 2012.

[18] David Pollard. *Convergence of Stochastic Processes*. Springer-Verlag, 1984.

[19] Hans Ulrich Simon. Bounds on the number of examples needed for learning functions. *SIAM Journal of Computing*, 26(3):751–763, 1997.

[20] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.

[21] Mathukumalli Vidyasagar. *A Theory of Learning and Generalization: With Applications to Neural Networks and Control Systems*. Springer-Verlag London Limited, London, 1997.

[22] Chao Zhang, Wei Bian, Dacheng Tao, and Weisi Lin. Discretized-vapnik-chervonenkis dimension for analyzing complexity of real function classes. *Neural Networks and Learning Systems, IEEE Transactions on*, 23(9):1461–1472, Sept 2012.

[23] Chao Zhang and Dacheng Tao. Structure of indicator function classes with finite vapnik-chervonenkis dimensions. *Neural Networks and Learning Systems, IEEE Transactions on*, 24(7):1156–1160, July 2013.