arXiv:2101.04043v1 [math.NA] 11 Jan 2021

Randomized weakly admissible meshes

Yiming Xu[1] and Akil Narayan[1,2]

[1]Department of Mathematics, University of Utah, yxu@math.utah.edu
[2]Scientific Computing and Imaging Institute, University of Utah, akil@sci.utah.edu

Abstract

A weakly admissible mesh (WAM) on a continuum real-valued domain is a sequence of discrete grids such that the discrete maximum norm of polynomials on the grid is comparable to the supremum norm of polynomials on the domain. The asymptotic rate of growth of the grid sizes and of the comparability constant must grow in a controlled manner. In this paper we generalize the notion of a WAM to a hierarchical subspaces of not necessarily polynomial functions, and we analyze particular strategies for random sampling as a technique for generating WAMs. Our main results show that WAM's and their stronger variant, admissible meshes, can be generated by random sampling, and our analysis provides concrete estimates for growth of both the meshes and the discrete-continuum comparability constants.

Keywords: Admissible meshes, Discrete meshes, Random sampling

1 Introduction

Generating a discrete set to approximate a continuum is a task that arises in many areas of applied computational science. A concrete example is that of computing a so-called weakly admissible (discrete) mesh for polynomials. Given a compact domain $D \subset \mathbb{R}^d$ and a fixed $n \in \mathbb{N}$, consider a discrete set $\mathcal{A}_n \subset D$ such that

$$\sup_{x\in D} |p(x)| \leq C_n \max_{x\in\mathcal{A}_n} |p(x)|\,, \qquad\qquad p \in P_n, \tag{1}$$

where P_n is the subspace of algebraic d-variate polynomials of degree n or less, and C_n is some finite constant that may depend on n. With $N_n \coloneqq \dim P_n$, any sequence of meshes $\{\mathcal{A}_n\}_{n=1}^{\infty}$ is called a *weakly admissible mesh* (WAM) [4] if there exist absolute constants $a, b, C, C' < \infty$ such that $|\mathcal{A}_n| \leq CN_n^a$ and $C_n \leq C'N_n^b$ for all $n \in \mathbb{N}$. It is an *admissible* mesh (AM) if $b = 0$ (see Definition 2.1 for more formal statements). An "optimal" WAM would boast the smallest values of these exponents, $a = 1$, $b = 0$, since large a implies increased growth of the mesh as n increases, and large b implies increased growth of the discrete-continuum equivalence constant as n grows.

Note that the reverse inequality is straightforward (with constants 1), so that a weakly admissible mesh yields equivalences between continuous and discrete supremum norms. This fact has been used to great effect in generating provably attractive meshes for polynomial approximation. However, it can be relatively difficult to generate such meshes for general domains D, and all existing constructions are essentially deterministic in nature.

The purpose of this article is to show that random sampling (i.e., comprised of independent and identically-distributed samples) can be used to generate meshes that are weakly admissible or admissible, and to prescribe a minimal number of required samples so that one can achieve inequalities of the form (1) with high probability, and with constants C_n whose large-n behavior is known. We also provide a straightforward generalization of WAM's to non-polynomial spaces, and all our results apply in this case. When d is "small", say 2 or 3, then one can use geometric analysis to construct deterministic grids that satisfy WAM properties [3]. When d is larger,

using such analysis becomes difficult, or at least the complexity of corresponding computational algorithms suffer the curse of dimensionality. However, in high-dimensional spaces one can frequently generate random samples from a domain without explicit knowledge of the geometry (e.g., rejection sampling or Markov Chain Monte Carlo methods). In some special cases there are algorithms that require only linear complexity in d to generate each sample. When it is feasible for meshes to be randomly generated, one supposes that a large enough number of samples can be used to form WAM's. We provide quantitative analysis for this procedure.

Our contributions in this article are twofold. We first generalize the notion of weakly admissible meshes to general hierarchical subspaces $\{V_n\}_{n=0}^{\infty}$, i.e., $V_n \subset V_{n+1}$ with $N_n := \dim(V_n)$. This generality allows us to consider generating meshes for approximation problems involving subspaces spanned by very general functions, not just polynomials. The general algorithmic procedure we consider in this article is as follows: With $\{\rho_n\}_{n=0}^{\infty}$ some sequence of probability measures, let $\mathcal{A}_n$ be generated randomly as $|\mathcal{A}_n|$ iid samples from ρ_n for each n. Our main results show that, for specific choices of ρ_n and $|\mathcal{A}_n|$, this procedure generates weakly admissible meshes. A summary of these results is as follows: Given a(ny) probability measure μ on $\mathbb{R}^n$ with closed support D such that $\{V_n\}_{n=0}^{\infty} \subset L^2_\mu(D)$, and any $\epsilon > 0$, the following statements are true with probability 1.

- Let $\rho_n \equiv \mu$ for all n. Then $|\mathcal{A}_n|$ can be chosen so that $a = q^* + \epsilon$ and $b = q^*/2 + \epsilon$, and hence such $\mathcal{A}_n$ form a WAM. See Theorem 3.4.
- Let $\rho_n = \mu_{V_n}$, where the latter is defined later in (5). Then $|\mathcal{A}_n|$ can be chosen so that $a = 1 + \epsilon$ and $b = q^*/2 + \frac{1}{2} + \epsilon$, and hence such $\mathcal{A}_n$ form a WAM. See Theorem 3.8.
- Assume D is convex and compact, and that the elements of V_n for each n are smooth. Let $\rho_n = \nu_n$, where the latter is defined later in (13). Then $|\mathcal{A}_n|$ can be chosen so that $a = p^* + \epsilon$ and $b = 0$, and hence such $\mathcal{A}_n$ form an AM. See Theorem 4.3.

Above, q^* is the asymptotic log-ratio of the Bernstein-Markov factor of the space V_n and N_n (see (9)), and p^* roughly measures the growth rate of certain weighted covering numbers of D relative to N_n (see (15)). Both q^* and p^* depend only on the prescribed measure μ and the hierarchical spaces V_n. The sampling measure μ_{V_n} introduced above was utilized in [7] to construct least squares approximations via random sampling. Due to the generality of our results, the first two results above are weaker than some existing results for polynomials, e.g., in [4]. We will elaborate on these statements in Section 2.

As a specialization of our hierarchical subspaces V_n, our methodology can be used to generate WAMs for polynomial spaces more exotic than the total degree spaces P_n. The lower bound on $|\mathcal{A}_n|$ to achieve (1) is $\dim P_n$, but many other hierarchical polynomial spaces have much smaller dimension. For example, the dimension of the subspace H_n containing d-dimensional functions spanned by homogeneous polynomials whose multi-index exponents lie in a hyperbolic cross index set of order n behaves like

$$\dim H_n \sim n \log^{d-1} n \ll \begin{pmatrix} n+d \\ d \end{pmatrix} = \dim P_n,$$

where the inequality is true for large n and $d > 1$ is large. Thus, an optimal discrete grid for a WAM has size comparable to $\dim H_n$, which can be significantly smaller than grids from a WAM generated for P_n.

The rest of this article is organized as follows. In Section 2, we introduce the notation and definitions. A few examples illustrating the definitions are provided to facilitate understanding. In Section 3, we borrow the idea of the near-isometry property of random matrices to obtain a sampling measure that generates weakly admissible meshes with exponents $a = q^* + \epsilon$ and $b = q^*/2 + \epsilon$, and then appropriately reweight the measure to make a near-optimal at mild cost on b: $a = 1 + \epsilon$ and $b = q^*/2 + 1/2 + \epsilon$. In Section 4, we introduce a novel sampling strategy to generate admissible meshes, and our analysis lies in a probabilistic argument using so-called weighted covering numbers. In Section 5, we provide numerical simulations to support our theoretical findings.

μ,D	probability measure on $\mathbb{R}^d$ and $D = \operatorname{supp}\mu$
V	N-dimensional subspace of functions in $L^2_\mu(D)$
μ_V	the V-induced measure on D
$K_\mu(V)$, $\lambda(x)$	the Bernstein-Markov factor for V in $L^2_\mu(D)$, and the normalized V-Christoffel function
R_n	square root of the sum of Christoffel functions induced by the partial derivatives
ν_n	the R_n-induced measure (weighted by R_n^d)
W	N-dimensional space of functions in V weighted by the Christoffel function
v_i, w_i	any orthonormal basis for V in L^2_μ, and for W in $L^2_{\mu_V}$, respectively
$\mathcal{A}$, M	A finite set in D of size M
$\boldsymbol{A}_{\mathcal{A}}$	$M \times N$ Vandermonde-like matrix associated with v_n on $\mathcal{A}$
$\|\cdot\|_{2,\mu}$, $\|\cdot\|_\infty$	$L^2_\mu(D)$ norm, and $L^\infty(D)$ norm, respectively
$\|\cdot\|_{2,\mathcal{A}}$, $\|\cdot\|_{\infty,\mathcal{A}}$	Discrete ℓ^2 and ℓ^∞ norms, respectively, on $\mathcal{A}$

Table 1: Notation used throughout this article.

2 Notation

Let $d \in \mathbb{N}$ be fixed. Consider a probability measure μ on $\mathbb{R}^d$ whose support is denoted $D \subseteq \mathbb{R}^d$. We do not require any particular conditions on D at present, but the results in Section 4 regarding admissible meshes require compactness of D. The Hilbert space $L^2_\mu(D;\mathbb{C})$ is endowed with the inner product and norm

$$\langle f, g\rangle_\mu \coloneqq \int_{\mathbb{R}^d} f(x)\bar{g}(x)\mathrm{d}\mu(x), \qquad\qquad \|f\|^2_{2,\mu} \coloneqq \langle f, f\rangle_\mu .$$

The supremum norm on D for functions f is

$$\|f\|_\infty = \sup_{x\in D} |f(x)| .$$

Note D may be unbounded. If $\mathcal{A} \subset D$ is a size-M set of points, we define discrete L^2 and L^∞ norms as

$$\|f\|^2_{2,\mathcal{A}} \coloneqq \frac{1}{M}\sum_{x\in\mathcal{A}} |f(x)|^2, \qquad\qquad \|f\|_{\infty,\mathcal{A}} \coloneqq \max_{x\in\mathcal{A}} |f(x)| .$$

Let V be an N-dimensional subspace of functions in $L^2_\mu(D)$. We can choose an associated orthonormal basis $v_1, \ldots, v_N \in V$. Let δ_x denote the x-centered Dirac delta distribution. For any $x \in D$, the V-valued function

$$K(x,\cdot) = \sum_{i=1}^N \bar{v}_i(x)v_i(\cdot),$$

is the V-Riesz representor of δ_x in L^2_μ. For any $v \in V$, we have

$$|v(x)| = \left|\langle v, K(x,\cdot)\rangle_\mu\right| \le \|v\|_{2,\mu}\,\|K(x,\cdot)\|_{2,\mu} = \|v\|_{2,\mu}\sqrt{\sum_{i=1}^N |v_i(x)|^2}.$$

And so,

$$\|v\|^2_{2,\mu} \le \|v\|^2_\infty \le K_\mu(V)\,\|v\|^2_{2,\mu}, \qquad\qquad K_\mu(V) \coloneqq \|K(x,x)\|_\infty . \tag{2}$$

where the lower inequality above holds since μ is a probability measure. Throughout this article we assume that $K_\mu(V)$ is finite, which excludes, e.g., polynomials on unbounded domains. The optimal (smallest) value of the equivalence factor $K_\mu(V)$ is N:

$$K_\mu(V) = \left\|\sum_{i=1}^N |v_i(x)|^2\right\|_\infty \ge \sum_{i=1}^N \|v_i(x)\|^2_{2,\mu} = N.$$

If V is chosen as a space of polynomials up to a certain degree, selecting μ as a so-called optimal measure achieves this optimal factor [1]. We also define

$$R(x) = \left(\sum_{i=1}^{N} \|\nabla v_i(x)\|^2 \right)^{1/2} \qquad R_\mu(V) = \|R(x)\|_\infty, \tag{3}$$

where $\|\cdot\|$ is the Euclidean norm on vectors. This function will play a role in our analysis involving covering numbers. In the following discussion, we assume that $R(x)$ is strictly positive on D.

2.1 Weighted spaces and induced probability measures

The L^2_μ-L^∞ equivalence established by (2) can be improved to the optimal equivalence if one considers weighted spaces: For a finite-dimensional V with L^2_μ-orthonormal basis v_i, define the (L^2) Christoffel function

$$\lambda(x) = \lambda_{V,\mu}(x) = \frac{N}{K(x,x)} = \frac{N}{\sum_{i=1}^{N} |v_i(x)|^2},$$

and consider the associated space of weighted elements from V:

$$W \coloneqq \sqrt{\lambda(x)} V \coloneqq \left\{ \sqrt{\lambda} v \mid v \in V \right\} \tag{4}$$

We also define a weighted measure μ_V via

$$\mathrm{d}\mu_V(x) \coloneqq \frac{1}{\lambda(x)} \mathrm{d}\mu(x) = \frac{1}{N} K(x,x) \mathrm{d}\mu(x), \tag{5}$$

which is another probability measure on D with $\mu_V \ll \mu$. The functions $w_i = v_i \sqrt{\lambda(x)}$ are an $L^2_{\mu_V}(D)$-orthonormal basis for W, and

$$K_{\mu_V}(W) = \left\| \sum_{i=1}^{N} |w_i(x)|^2 \right\|_\infty = N,$$

so that μ_V is an optimal measure for W, and we have the optimal equivalence relation

$$\|w\|_{2,\mu_V} \leq \|w\|_\infty \leq N \|w\|_{2,\mu_V}, \qquad w \in W \tag{6}$$

The measure μ_V has utility in recent computational strategies for constructing discrete least-squares approximations [7]. In this article, we will call μ_V the V-induced measure for μ. The term "induced" stems from historical context: For certain μ and polynomial spaces V, the measure μ_V is an additive mixture of tensor-product measures; the univariate measures that define the tensor-product measures in this case are similar to induced orthogonal polynomials [9]. Sampling from such non-standard measures is computationally efficient and feasible by exploiting properties of orthogonal polynomials [13].

2.2 Polynomial spaces

We will sometimes be concerned with the special case when V is a subspace of polynomials. In this specialized case we denote the space P as an N-dimensional space of polynomials. It is convenient (but not necessary) to use multi-indices to define these spaces.

We let $\alpha \in \mathbb{N}_0^d$ denote a d-dimensional multi-index and use $\Sigma \subset \mathbb{N}_0^d$ to denote a finite set of multi-indices. Associated to any Σ, we define the subpsace of algebraic polynomials spanned by monomials:

$$P_\Sigma \coloneqq \operatorname{span} \{ x^\alpha \mid \alpha \in \Sigma \} \tag{7}$$

A particularly special set of multi-indices are those corresponding to the total-degree space of polynomials:

$$\Sigma_n = \left\{\alpha \in \mathbb{N}_0^d \mid |\alpha| \leq n\right\}, \qquad\qquad n \in \mathbb{N}_0.$$

We will use the abbreviation $P_n \coloneqq P_{\Sigma_n}$.

Finally, we note that there are many finite-dimensional polynomial spaces that cannot be written in the form (7). This is a deficiency in our presentation style in that we emphasize the specific class of subspaces (7). However, all our theoretical results extend to general polynomial subspaces.

2.3 WAM's for general hierarchical subspaces

We generalize the notion of admissibility to general hierarchical subspaces. Let $\{V_n\}_{n=0}^{\infty}$ denote any sequence of finite-dimensional hierarchical subspaces, i.e., $V_n \subset V_{n+1}$ and $\dim V_n < \dim V_{n+1}$ for all n. We set

$$N_n \coloneqq \dim V_n,$$

which is a sequence of strictly increasing positive integers.

Definition 2.1 (Weakly admissible meshes for hierarchical spaces)**.** *Let D be a closed set in $\mathbb{R}^d$. Consider $\{\mathcal{A}_n\}_{n=0}^{\infty}$ a collection of finite subsets of D. Assume there is a collection of constants C_n, $n \geq 0$, such that*

- $\|v\|_\infty \leq C_n \|v\|_{\infty,\mathcal{A}_n}$ *for all* $v \in V_n$
- $C_n = \mathcal{O}\left(N_n^b\right)$ *for some* $b < \infty$.
- $|\mathcal{A}_n| = \mathcal{O}\left(N_n^a\right)$ *for some* $a < \infty$.

Then $\{\mathcal{A}_n\}_{n=0}^{\infty}$ is called a weakly admissible mesh (WAM). It is called an admissible mesh (AM) if $b = 0$.

Setting $V_n = P_n$, our definition above is consistent with the one used in [3]. The optimal value of the exponent b is $b = 0$, and such meshes are known to exist for domains exhibiting a polynomial Markov inequality [4], although the construction relies on grids that achieve certain fill distances and can thus be cumbersome for sufficiently complex domains using a deterministic approach. The optimal value of the exponent a in general is $a = 1$ since the inequality

$$\|p\|_\infty \leq C_n \|p\|_{\infty,\mathcal{A}_n}, \qquad\qquad p \in V_n \tag{8}$$

holds for some finite C_n only if $\mathcal{A}$ is determining for P_n (i.e., for any $p \in V_n$, $p(x) = 0$ for all $x \in \mathcal{A}$ implies $p \equiv 0$).

Note that in the modified definition D is allowed to be unbounded, meaning that WAM's for V_n are only sensible if V_n contains functions that are bounded on D. This subtlety will be reiterated when we discuss random sampling for generating admissible meshes. In the remainder of this paper, we will refer to the above definition when speaking of a WAM.

We will make a further assumption on the spaces V_n that involves the L^2_μ machinery we have introduced, namely that they satisfy

$$q^* \coloneqq q^*\left(\mu, \{V_n\}_{n=0}^{\infty}\right) = \limsup_{n\to\infty} \frac{\log K_\mu(V_n)}{\log N_n} < \infty. \tag{9}$$

Note that since $K_\mu(V_n) \geq \dim V_n = N_n$, then $q^* \geq 1$. Our main results using concentration of measure characterize the WAM exponents a and b via proportionality to q^*, and thus small q^* is desirable. Requiring finite q^* can be related to similar notions in the polynomial context. If we choose $V_n = P_n$, then finite q^* along with

$$|\mathcal{A}| \geq N_n = \dim P_n = \binom{n+d}{d} \sim \frac{n^d}{d!} \tag{10}$$

and the fact the $N_n^{1/n} \to 1$ implies

$$\lim_{n\to\infty} K_\mu(P_n)^{1/n} = 1,$$

showing that the pair (D, μ) satisfies the so-called Bernstein-Markov property. Pairs that satisfy the Bernstein-Markov property have fundamental connections to various results in approximation theory. (E.g., section 5 of [2] for a summary.) In fact, if μ satisfies a "density condition" then it is known that q^* is finite for sequences of fairly general polynomial subspaces [11, Corollary 4.2.2]. Thus, our requirement that $q^* < \infty$ is not unnatural, but is slightly stronger than a Bernstein-Markov property when specialized to polynomials.

To illustrate values of q^*, we summarize three special choices for μ and V_n.

Example 2.1 Complex exponentials
Let $\mathrm{d}\mu(x) = \mathrm{d}x$ on the unit cube $D = [0,1]^d$ for arbitrary $d \geq 1$. For an arbitrary subset $F \subset \mathbb{Z}^d$ of size N given by $F = \{f_1, \ldots, f_N\}$, define

$$v_n(x) = \exp\left[2\pi i (f_n \cdot x)\right], \qquad V = \operatorname{span}\left\{v_n\right\}_{n=1}^N,$$

where $f_n \cdot x$ is the standard componentwise inner product between two elements in $\mathbb{R}^d$. Then v_n is an orthonormal basis for V, and $K_\mu(V) = N$, and thus we can choose any hierarchical collection of subspaces $V_n \subset V_{n+1}$ defined by corresponding hierarchical sets $F_n \subset F_{n+1}$. We then have

$$q^* = \limsup_{n\to\infty} \frac{\log K_\mu(V_n)}{\log N_n} = \limsup_{n\to\infty} \frac{\log N_n}{\log N_n} = 1,$$

thus achieving the optimal q^* factor. In this case we also have $\mu_V = \mu$. In the language of [1] for polynomials, the measure μ is an optimal measure.

Example 2.2 Tensor-product Jacobi polynomials
Let $\mathrm{d}\mu(x) \propto \prod_{j=1}^d \left(1 - x^{(j)}\right)^\alpha \left(1 - x^{(j)}\right)^\beta$ on $x \in [-1,1]^d = D$ and $\alpha, \beta \in \mathbb{N}_0$. Choosing $V_n = P_n$, the space of d-variate polynomials of degree n or less, an orthonormal family for V_n is provided by tensorized Jacobi polynomials. The estimates in [12] show that $K_\mu(V_n) \leq N_n^{2(\gamma+1)}$, where $\gamma = \max\{\alpha, \beta\}$. Therefore,

$$q^* = \limsup_{n\to\infty} \frac{\log K_\mu(V_n)}{\log N_n} \leq \limsup_{n\to\infty} \frac{2(\gamma+1)\log N_n}{\log N_n} = 2(\gamma+1).$$

Here, while N_n and $K_\mu(V_n)$ both grow exponentially in d, the quantity q^* does not.

Example 2.3 Tensor-product Chebyshev polynomials
With μ and V_n as in the previous example, now take $\alpha = \beta = -1/2$, so that an orthonormal basis is provided by tensorized Chebyshev polynomials. Univariate Chebyshev polynomials $T_k(y)$, $y \in [-1,1]$, satisfy $T_k^2(y) \leq 2$ for all k, so that $K_\mu(V_n) \leq N_n 2^d$. Thus,

$$q^* = \limsup_{n\to\infty} \frac{\log K_\mu(V_n)}{\log N_n} \leq \limsup_{n\to\infty} \frac{d\log 2 + \log N_n}{\log N_n} = 1.$$

Since $q^* \geq 1$ always holds, we conclude that $q^* = 1$. Note that the bound $K_\mu(V_n) \leq N_n 2^d$ holds when $V_n = P_\Sigma$ for any multi-index set Σ. Thus, the behavior $q^* = 1$ holds for very general hierarchical polynomial spaces. This suggests that the Chebyshev measure μ is n-asymptotically optimal.

2.4 Weighted covering

Designing an AM, a stronger WAM with $b = 0$, via random sampling relies on generating grids with good space-filling properties. A baseline is the uniform sampler, and the corresponding sufficient sampling size can be obtained by analyzing the covering number of the domain. In

the deterministic context, low discrepancy sequences [8] are a standard approach for "uniformly" filling a volume with points. However, this approach is more difficult when D is not a hypercube, and so we will investigate randomized approaches via sampling. To understand how points generated from an arbitrary measure fill the domain, we introduce the following definition of weighted covering:

Definition 2.2 (f-weighted covering)**.** *Let $D \subset \mathbb{R}^d$ be compact and $f : D \to \mathbb{R}_+$ be a continuous function, where $\mathbb{R}_+ = (0, \infty)$. For $r > 0$, let $B_r(x) = \{z \in \mathbb{R}^d : \|z - x\|_2 \leq r\}$. Let $m_f = \min_{u \in D} f(u) > 0$. For $y \in D$ and $r > 0$, define*

$$F_r(y) = \frac{\min_{z \in B_r(y) \cap D} f(z)}{m_f}.$$

A set $\mathcal{N} \subset D$ is called an f-weighted ϵ-covering of D if

$$D \subset \bigcup_{y \in \mathcal{N}} B_{r(y,\epsilon)}(y) \qquad\qquad r(y, \epsilon) = \max_{c \geq 0} \min \{\epsilon F_c(y), c\}. \tag{11}$$

In particular, when f is a constant function, an f-weighted ϵ-covering is the same as an ϵ-covering.

The definition of $r(y, \epsilon)$ may look mysterious at first sight: The general idea behind the f-weighted ϵ-covering is to cover points in D using balls of radius proportional to f, with the location(s) $\operatorname{argmin}_{u \in D} f(u)$ covered by balls of radius ϵ. However, under this definition points within the same ball in a covering may have very different weights attached to them. To address this, we require that the ratio between weight of any point in the same ball in a covering and m_f must be at least r/ϵ, where r the radius of the ball. Simultaneously, we wish to choose r as large as possible (hence the max-min condition in (11)).

Since $F_r(y) \geq 1$ for any $y \in D$ and $r > 0$, then $r(y, \epsilon) \geq \epsilon$. This implies that any ϵ-covering is also an f-weighted epsilon covering. Also, for fixed ϵ, $F_c(y)$ is a continuous and non-increasing function of c which evaluates to 1 for $c \geq \operatorname{diam}(D)$. Therefore,

$$r(y, \epsilon) = \epsilon F_{r(y,\epsilon)}(y). \tag{12}$$

It follows easily from (12) that $r(y, \epsilon)$ is non-increasing in ϵ. As an immediate consequence, for $0 < \epsilon_1 < \epsilon_2 \leq 1$,

$$\frac{r(y, \epsilon_1)}{r(y, \epsilon_2)} = \frac{\epsilon_1 F_{r(y,\epsilon_1)}(y)}{\epsilon_2 F_{r(y,\epsilon_2)}(y)} \geq \frac{\epsilon_1}{\epsilon_2}.$$

This implies that the radius of a weighted covering ball at a given point scales slower than ϵ, leading to a potentially better constant for the rate of covering number as $\epsilon \to 0$, which is a property that we exploit.

Denote $R_n(x)$ as the function in (3) when $V = V_n$ and choose $f(x) = R_n(x)^{-1}$. Since $R_n(x)$ is positive on D, $f(x)$ will satisfy the assumption in Definition 2.2 if $R_n(x)$ is continuous. The f-weighted covering number will be used to analyze the sampling properties of the R_n-weighted probability measure on D which is defined as

$$d\nu_n(x) = \frac{R_n^d(x)}{\int_D R_n^d(x) dx} dx \qquad\qquad x \in D. \tag{13}$$

Note that (13) is well-defined as long as $R_n(x) \in L^d(D)$. The measure $\nu_n(x)$ plays a crucial role in the design of sampling for admissible meshes. Let $\mathcal{S}_{f,\epsilon}(D)$ be the set of f-weighted ϵ-coverings of D. For any $\mathcal{N} \in \mathcal{S}_{f,\epsilon}(D)$, define

$$G_{\mathcal{N}} = \sup_{y \in \mathcal{N}} \left(\frac{\max_{x \in B_{r(y,\epsilon)}} f(x)}{\min_{x \in B_{r(y,\epsilon)}} f(x)} \right)^d, \tag{14}$$

which evaluates how "conservative" the f-weighted covering $\mathcal{N}$ is. A quantity that will appear in our analysis for the AM exponents is

$$p^* = \limsup_{n\to\infty} \frac{\log\left(\inf_{\mathcal{N}\in\mathcal{S}_{f,\epsilon}(D)} |\mathcal{N}| G_{\mathcal{N}}\right)}{\log N_n} \qquad \epsilon = (3R_\mu(V_n))^{-1}, \tag{15}$$

which we assume to be finite. The constant 3 is non-essential and can be replaced with any constant greater than 2. Similar to q^* for WAM's, p^* will characterize the exponents for AM in our result. Note that p^* is mostly of theoretical interest: Explicit estimation of p^* is essentially impossible unless the underlying geometry of the domain is simple. For example, for Chebyshev polynomials with $d = 1$, one can numerically verify that $p^* \leq 1.70$ (see Section 5).

3 Randomized weakly admissible meshes

3.1 Discrete randomized near-isometries

The main strategy for our approach comes in two parts: First we generate a finite mesh $\mathcal{A}$ that emulates the L^2_μ norm on V. We subsequently use that mesh and the L^2-L^∞ equivalence relations described earlier in order to transform comparability of $\|\cdot\|_{2,\mu}$ and $\|\cdot\|_{2,\mathcal{A}}$ into comparability between $\|\cdot\|_\infty$ and $\|\cdot\|_{\infty,\mathcal{A}}$. The first part of this strategy, the construction of $\mathcal{A}$ based on L^2_μ properties, is the subject of this section.

With $\{v_i\}_{i=1}^N$ an orthonormal basis for V, we require the following algebraic formulation: the $M \times N$ matrix $\boldsymbol{A}_{\mathcal{A}}$ has entries

$$(\boldsymbol{A}_{\mathcal{A}})_{m,i} = \frac{1}{\sqrt{M}} v_i(x_m), \qquad \mathcal{A} = \{x_m\}_{m=1}^M, \quad (m,i) \in [M]\times[N].$$

The problem of finding a discrete mesh capable of emulating the L^2_μ norm is conceptually identical to finding a stable discrete least-squares problem defined by the matrix $\boldsymbol{A}$. We codify this in the following theorem from [6] with a more explicit constant.

Theorem 3.1 ([6]). *Let $\mathcal{A} = \{X_m\}_{m=1}^M$, where the X_m are independent and identically distributed draws of a random variable distributed according to the probability measure μ. For any $r > 0$ and $0 < \delta < 1$, suppose that*

$$\frac{M}{\log M} \geq \frac{3(1+r)}{\delta^2} K_\mu(V). \tag{16}$$

Then

$$\Pr\left[\left\|\boldsymbol{A}_{\mathcal{A}}^T \boldsymbol{A}_{\mathcal{A}} - \boldsymbol{I}\right\| \geq \delta\right] \leq 2M^{-r}, \tag{17}$$

where $\|\cdot\|$ is the induced (spectral) norm on matrices.

The quantity $\boldsymbol{A}_{\mathcal{A}}^T \boldsymbol{A}_{\mathcal{A}}$ is the Gramian of the basis v_i with respect to the discrete inner product $\langle\cdot,\cdot\rangle_{2,\mathcal{A}}$, and thus (17) quantifies the L^2 proximity of a discrete measure supported on $\mathcal{A}$ to μ on the space V. The sample count complexity (16) couples M and N with dependence on the proximity parameter δ and the success parameter r, but also involves the Bernstein-Markov factor $K_\mu(V)$. When V is a polynomial space defined by multi-index set Σ, many standard continuous probability measures yield extremely large $K_\mu(V)$, often depending exponentially on d and algebraically on the maximum polynomial degree in Σ [5].

The authors in [10, 14] note that introducing weights related to $\lambda(x)$ into the least-squares algorithm would result in a procedure with optimal (minimal) sample count. The authors in [7] propose a computationally feasible procedure for drawing samples from the induced measure μ_V using this weighting idea, and arrive at the following result:

Theorem 3.2 ([7]). *Let $\mathcal{A} = \{X_m\}_{m=1}^M$, where the X_m are independent and identically distributed draws of a random variable distributed according to the probability measure μ_V. Introduce weights $\omega_m := \lambda_V(x)$, and define the diagonal $M \times M$ matrix $\boldsymbol{W}$ with entries $(W)_{m,m} = \omega_m$. With r and δ as in Theorem 3.1, assume*

$$\frac{M}{\log M} \geq \frac{3(1+r)}{\delta^2} N. \tag{18}$$

Then

$$\Pr\left[\left\|\boldsymbol{A}_{\mathcal{A}}^T \boldsymbol{W} \boldsymbol{A}_{\mathcal{A}} - \boldsymbol{I}\right\| \geq \delta\right] \leq 2M^{-r}. \tag{19}$$

Note that the sample complexity (18) is near-optimal, up to the $\log M$ factor.

If a mesh $\mathcal{A}$ satisfies the conditions of either Theorem 3.1 or 3.2, then we can establish equivalences between discrete and continuous L^2 norms.

Corollary 3.1. *1. Suppose $\mathcal{A}$ is a mesh satisfying the conditions of Theorem 3.1. Then with probability at least $1 - 2M^{-r}$,*

$$\frac{1}{1+\delta} \|v\|_{2,\mathcal{A}}^2 \leq \|v\|_{2,\mu}^2 \leq \frac{1}{1-\delta} \|v\|_{2,\mathcal{A}}^2, \qquad v \in V \tag{20a}$$

2. Let $\mathcal{A}$ be a mesh satisfying the conditions of Theorem 3.2. Then with probability at least $1 - 2M^{-r}$,

$$\frac{1}{1+\delta} \|w\|_{2,\mathcal{A}}^2 \leq \|w\|_{2,\mu_V}^2 \leq \frac{1}{1-\delta} \|w\|_{2,\mathcal{A}}^2, \qquad w \in W, \tag{20b}$$

where W is the space (4) of $\sqrt{\lambda}$-weighted V functions.

Proof. We prove the second statement; the proof of the first statement is similar. For an arbitrary $w \in W$, represent $w(x) = \sum_{i=1}^N u_i w_i(x)$ according to the $L^2_{\mu_V}$-orthonormal basis $\{w_i\}_{i=1}^N$, so that

$$\|w\|_{2,\mu_V} = \|\boldsymbol{u}\|,$$

for $\|\cdot\|$ the Euclidean norm on vectors. Also,

$$\left(\sqrt{\boldsymbol{W}} \boldsymbol{A}_{\mathcal{A}} \boldsymbol{u}\right)_m = \sum_{i=1}^N \sqrt{\frac{\lambda(X_m)}{M}} u_i w_i(X_m) = \frac{1}{\sqrt{M}} \sum_{i=1}^N u_i w_i(X_m) = \frac{1}{\sqrt{M}} w(X_m).$$

Thus, $\|w\|_{2,\mathcal{A}} = \left\|\sqrt{\boldsymbol{W}} \boldsymbol{A}_{\mathcal{A}} \boldsymbol{u}\right\|$. Assuming the complement of the probabilistic event in (19),

$$\begin{aligned} \|w\|_{2,\mathcal{A}}^2 = \left\|\sqrt{\boldsymbol{W}} \boldsymbol{A}_{\mathcal{A}} \boldsymbol{u}\right\|^2 &= \boldsymbol{u}^T \left(\boldsymbol{I} + \left(\boldsymbol{A}_{\mathcal{A}}^T \boldsymbol{W} \boldsymbol{A}_{\mathcal{A}} - \boldsymbol{I}\right)\right) \boldsymbol{u} \\ &\leq \|\boldsymbol{u}\|^2 (1+\delta) = (1+\delta) \|w\|_{2,\mu_V}^2. \end{aligned}$$

This establishes the lower inequality in (20b). The upper inequality is shown in the same way:

$$\begin{aligned} \|w\|_{2,\mathcal{A}}^2 = \left\|\sqrt{\boldsymbol{W}} \boldsymbol{A}_{\mathcal{A}} \boldsymbol{u}\right\|^2 &= \boldsymbol{u}^T \left(\boldsymbol{I} - \left(\boldsymbol{I} - \boldsymbol{A}_{\mathcal{A}}^T \boldsymbol{W} \boldsymbol{A}_{\mathcal{A}}\right)\right) \boldsymbol{u} \\ &\geq \|\boldsymbol{u}\|^2 (1-\delta) = (1-\delta) \|w\|_{2,\mu_V}^2. \end{aligned}$$

□

3.2 Sampling from μ

Since V is a subspace of L^2_μ, it is reasonable to believe that taking a large number of iid samples from μ will eventually allow one to approximate the L^∞ norm of any element in V. Theorem 3.3 below shows that for a fixed subspace V, iid samples yield an equivalence relation between the discrete and continuous maximum norms.

Theorem 3.3. *Let V be a given subspace of dimension N, and assume that M is large enough to satisfy (16) for some $\delta \in (0,1)$ and $r > 0$. Let $\mathcal{A}$ have size M with elements comprised of iid samples from μ. Then, with probability $1 - 2M^{-r}$, we have*

$$\|v\|_\infty \le \sqrt{\frac{K_\mu(V)}{1-\delta}}\, \|v\|_{\infty,\mathcal{A}}\,, \qquad\qquad v \in V. \tag{21}$$

Proof. For any $v \in V$, we have

$$\|v\|_\infty^2 \overset{(2)}{\le} K_\mu(V)\, \|v\|_{2,\mu}^2 \overset{(20a)}{\le} \frac{K_\mu(V)}{1-\delta}\, \|v\|_{2,\mathcal{A}}^2 \le \frac{K_\mu(V)}{1-\delta}\, \|v\|_{\infty,\mathcal{A}}^2\,, \tag{22}$$

where the second inequality holds with probability $1 - 2M^{-r}$. □

Note that this theorem appears quite suboptimal: not only do we require M/N to scale like $K_\mu(V)$, but we also pay a penalty factor of $K_\mu(V)$ in the norm comparability result (21). Nevertheless, we can use this construction to form weakly admissible meshes: Theorem 3.3 together with the Borel-Cantelli lemma yields the following result.

Theorem 3.4. *Let $\{V_n\}_{n=1}^\infty$ be given. For each n, define*

$$\mathcal{A}_n \coloneqq \{X_m\}_{m=1}^{M_n}\,, \qquad\qquad X_m \sim \mu.$$

Assume that $M_n = cK_\mu(V_n)\log N_n$ for some $c > 25q^*$. Then, with probability 1, $\{\mathcal{A}_n\}_{n=1}^\infty$ forms a WAM for V_n and μ with exponents $a = q^* + \tau$ for any $\tau > 0$, and $b = q^*/2 + \tau$.

Proof. We use the abbreviation $K_n \coloneqq K_\mu(V_n)$ to reduce notational clutter. Set $\delta = 1/2$. Our assumptions ensure that

$$\begin{aligned}\liminf_{n\to\infty} \frac{M_n}{K_n \log M_n} &= \liminf_{n\to\infty} \frac{c\log N_n}{\log K_n + \log(c\log N_n)} \\ &= \liminf_{n\to\infty} \frac{c\log N_n}{\log K_n} \overset{(9)}{=} \frac{c}{q^*} > 25 > 12(1+r)\end{aligned}$$

for $1 < r < 13/12$, so that for sufficiently large n,

$$\frac{M_n}{\log M_n} \ge 12(1+r)K_n = \frac{3(1+r)}{\delta^2}K_n.$$

The above condition verifies that (16) is satisfied. It follows from Theorem 3.3 that for any $\tau > 0$ and $1 < r < 13/12$, with probability $1 - M_n^{-r}$,

$$\|v\|_\infty \le \sqrt{2K_\mu(V)}\, \|v\|_{\infty,\mathcal{A}} \le C(\tau) N_n^{q^*/2+\tau}\, \|v\|_{\infty,\mathcal{A}}\,, \qquad\qquad v \in V, \tag{23}$$

where $C(\tau) > 0$ is some constant depending only on τ. Define E_n as the probabilistic event that the above inequality holds. Since

$$\sum_{n=1}^\infty \Pr(E_n^c) \le 2\sum_{n=1}^\infty M_n^{-r} \le 2\sum_{n=1}^\infty n^{-r} < \infty,$$

by the Borel-Cantelli Lemma, the probability that E_n^c happens infinitely often is 0, i.e., for any realization of the $\{\mathcal{A}_n\}_{n=1}^\infty$, then with probability 1, (23) holds for all n sufficiently large. The proof is complete. □

One can rephrase Theorem 3.4 as a probabilistic statement to obtain finer control on values of n that lie in the asymptotic regime. However, due to the lim sup effect it is difficult to control this for every n.

3.3 Sampling from μ_V

Meshes generated by randomly sampling from μ are suboptimal, as shown above. The WAM exponents of such meshes are effectively $a = q^*$ and $b = q^*/2$. We can entirely remove the dependence on q^* by considering weighted meshes. This section essentially repeats the computations of the previous section, but by replacing μ with μ_V and V with W. Since the proofs are almost identical to the ones in the previous section, we omit them for brevity.

Theorem 3.5. *With V given, let $\{X_m\}_{m=1}^M$ be a sequence of iid random variables distributed according to μ_V. Assume that M satisfies (18) for some $r > 0$ and $\delta < 1$. Then with probability at least $1 - 2M^{-r}$,*

$$\|w\|_\infty \le \sqrt{\frac{N}{1-\delta}} \|w\|_{\mathcal{A},\infty}, \qquad w \in W \tag{24}$$

The above shows that a sequence of randomizes grids as constructed above form a weakly admissible mesh.

Theorem 3.6. *Let $\{V_n\}_{n=1}^\infty$ be given. For each V_n, define*

$$\mathcal{A}_n \coloneqq \{X_{n,m}\}_{m=1}^{M_n}, \qquad X_{n,m} \sim \mu_{V_n}.$$

Assume that $M_n = 25N_n \log N_n$. Then $\{\mathcal{A}_n\}_{n=1}^\infty$ forms a weakly admissible mesh for W_n with exponents $a = 1 + \tau$ for any $\tau > 0$, and $b = \frac{1}{2}$.

The WAM's described above provide ways to bound supremum norms of functions in the weighted space W. We can translate these back into estimates on the space V by paying a mild penalty factor.

Theorem 3.7. *Let μ, D, and V be given, and assume that $1 \in V$. Let $\{X_m\}_{m=1}^M$ be iid samples from μ_V. If, for some $r > 0$ and $0 < \delta < 1$, M is large enough to satisfy (16), then with probability at least $1 - 2M^{-r}$,*

$$\|v\|_\infty \le \sqrt{N}\sqrt{\frac{K_\mu(V)}{1-\delta}} \|v\|_{\mathcal{A},\infty}, \qquad v \in V.$$

Proof. Since V contains constant functions, then let $v_1 = 1$ be a particular choice of the first element in an orthonormal basis for V. Then we have

$$\lambda(x) = \frac{N}{\sum_{n=1}^N v_n^2} \le N.$$

Any $w \in W$ can be written as $\sqrt{\lambda}v$ for some $v \in V$, and so

$$\|w\|_\infty \ge \|v\|_\infty \inf_{x \in D} \lambda(x) \ge \|v\|_\infty \sqrt{\frac{N}{K_\mu(V)}} \tag{25a}$$

Likewise,

$$\|w\|_{\infty,\mathcal{A}} \le \left\|\sqrt{\lambda}\right\|_{\infty,\mathcal{A}} \|v\|_{\infty,\mathcal{A}} \le \sqrt{N} \|v\|_{\infty,\mathcal{A}}. \tag{25b}$$

Chaining (24) with relations (25) proves the theorem. □

Finally, we can generate a WAM using the result above:

Theorem 3.8. *Let μ, D, and $\{V_n\}_{n=1}^\infty$ be given, and assume that $1 \in V_n$ for some n. For each n, define*

$$\mathcal{A}_n \coloneqq \{X_{n,m}\}_{m=1}^{M_n}, \qquad X_{n,m} \sim \mu_{V_n}.$$

Assume that $M_n = 25N_n \log N_n$. Then $\{\mathcal{A}_n\}_{n=1}^\infty$ forms a weakly admissible mesh for V_n with exponents $a = 1 + \tau$ for any $\tau > 0$, and $b = \frac{q^*+1}{2}$.

4 Randomized admissible meshes

The previous section focuses on devising sampling strategies to obtain an optimal sampling size at the expense of rendering the equivalence coefficient C_n being reasonably large. Alternatively, one may expect $C_n = \mathcal{O}(1)$ when the sampling size is sufficient, so that randomly generated grids almost fill the domain. This is similar to the idea in [4] but avoids deterministically discretizing domains with complicated geometry. With random samples, such a result is not possible using our previous analysis. One expects that attaining an admissible mesh from random samples is asymptotically possible (i.e., as the sample count increases to infinity) for all sampling measures that are absolutely continuous with respect to the uniform measure on D. However, in terms of finite-sample behavior, different measures may demonstrate drastically unequal performance. We will begin our discussion by taking the uniform measure as the baseline, from which we will learn how to design a sampling measure that achieves better efficiency. Since our analysis is based on a covering argument, we assume in the following that D is both convex and compact.

4.1 Sampling from the uniform measure

Suppose that $\mathcal{A}_n$ is a set of points independently and uniformly sampled from D. The next theorem shows that when M_n is logarithmically larger than the cardinality of some optimal covering of D, with overwhelming probability, $b = 0$.

Theorem 4.1. *Given μ, D, and $\{V_n\}_{n=1}^{\infty}$, define*

$$\mathcal{A}_n \coloneqq \{X_m\}_{m=1}^{M_n} \qquad\qquad X_m \sim \mathsf{Unif}(D).$$

Assume that D is compact and convex and elements in V_n are twice continuously differentiable. Fix $k > 2$ and $r > 1$. Let $L_n = \max\{N_n, |\mathcal{N}_n|\}$, where $\mathcal{N}_n \subset D$ is a $(kR_\mu(V_n))^{-1}$-covering of D. If $M_n \geq (r+1)L_n \log L_n$, then with probability at least $1 - N_n^{-r}$,

$$\|v\|_\infty \leq \frac{k}{k-2}\|v\|_{\mathcal{A}_n,\infty}.$$

Proof. Take

$$v(x) = \sum_{i\in[N_n]} \langle v, v_i\rangle_\mu v_i(x) \in V_n.$$

Without loss of generality assume $\|v\|_{2,\mu} = \sum_{i\in[N_n]}\langle v, v_i\rangle_\mu^2 = 1$, otherwise consider $v/\|v\|_{2,\mu}$. For $x, y \in D$, the segment connecting x and y is in D under the convexity assumption. Set $L = \|y - x\|_2$ and $z = L^{-1}(y - x)$. Applying the Fundamental Theorem of Calculus, we have

$$\begin{aligned}
|v(y) - v(x)| &= \left|\int_0^L \langle \nabla v(x + tz), z\rangle dt\right| && (26)\\
&= \left|\sum_{i\in[N_n]} \langle v, v_i\rangle_\mu \int_0^L \langle \nabla v_i(x + tz), z\rangle dt\right| \\
&\leq \left(\sum_{i\in[N_n]} \left|\int_0^L \langle \nabla v_i(x + tz), z\rangle dt\right|^2\right)^{1/2} \\
&\leq \left(\sum_{i\in[N_n]} L\int_0^L |\langle \nabla v_i(x + tz), z\rangle|^2 \, dt\right)^{1/2} \\
&\leq \left(L\int_0^L \sum_{i\in[N_n]} \|\nabla v_i(x + tz)\|^2 \, dt\right)^{1/2} \leq LR_\mu(V_n), && (27)
\end{aligned}$$

which implies that $|v(x) - v(y)| < \epsilon$ if $\|x - y\|_2 < R_\mu(V_n)^{-1}\epsilon$. Take $\epsilon = 1/k$. To reduce clutter, let

$$\eta = (kR_\mu(V_n))^{-1}. \tag{28}$$

Suppose that $\mathcal{N}_n \subset D$ is an η-covering of D with $|\mathcal{N}_n| \geq N_n$, and $\mathcal{A}_n \subset D$ satisfies

$$\mathcal{A}_n \cap B_\eta(y) \neq \varnothing \qquad\qquad \forall y \in \mathcal{A}_n. \tag{29}$$

It follows from the triangle inequality that for every element in D one can find a point in $\mathcal{A}_n$ such that their distance is at most 2η. Therefore,

$$\frac{\|v\|_\infty}{\|v\|_{\mathcal{A}_n,\infty}} \overset{k>2}{\leq} \frac{\|v\|_\infty}{\|v\|_\infty - \frac{2}{k}} \overset{\|v\|_\infty \geq \|v\|_{2,\mu}=1}{\leq} \frac{k}{k-2}. \tag{30}$$

We now show that for sufficiently large M_n, with high probability $\mathcal{A}_n$ satisfies (29). In fact, for $y \in \mathcal{N}_n$,

$$\begin{aligned} &\Pr\left[B_\eta(y) \cap \{X_m\}_{m\in[M_n]} = \varnothing\right] \\ =\ &\left(\Pr\left[B_\eta(y) \cap X_1 = \varnothing\right]\right)^{M_n} \leq \left(1 - \frac{1}{|\mathcal{N}_n|}\right)^{M_n} \leq e^{-\frac{M_n}{|\mathcal{N}_n|}}. \end{aligned}$$

Taking a union bound over y yields that the probability of $\mathcal{A}_n$ not satisfying (29) is at most $|\mathcal{N}_n| e^{-\frac{M_n}{|\mathcal{N}_n|}}$. Setting $M_n = (r+1)|\mathcal{N}_n| \log |\mathcal{N}_n|$ completes the proof. □

Remark 4.1. To get an idea of how large M_n grows as $n \to \infty$, take $D = B_1(0) \subset \mathbb{R}^d$. In this case,

$$L_n \leq (3kR_\mu(V_n))^d,$$

so that $M_n \geq (r+1)d(3kR_\mu(V_n)^d \log(3kR_\mu(V_n))$ ensures the result in Theorem 4.1 holds, i.e., $M_n = \mathcal{O}(dR_\mu(V_n)^d \log R_\mu(V_n))$.

4.2 Sampling from ν_n

Note that in the derivation of (27) the integrand is directly bounded by $R_\mu(V_n)$, which is a global quantity of V_n. We now propose an alternative random sampling strategy which aims to exploit the local structures of V_n so that the analysis of covering becomes more efficient. Precisely, we wish to cover points with large gradients using smaller balls. Keeping the steps before the last step in (27) yields

$$|v(y) - v(x)| \leq L \underbrace{\left(\frac{1}{L}\int_0^L R_n^2(x+tz)dt\right)^{\frac{1}{2}}}_{(*)}. \tag{31}$$

When L is small, $(*) \approx R_n(x)$, so the local Lipschitz constant of v at x is bounded by $R_n(x)$. In other words, evaluation of v at points within $R_n(x)^{-1}\epsilon$ distance to x change from $v(x)$ by approximately ϵ. Based on this observation, we propose an alternative sampling strategy making use of such local information, resulting in the Theorem below.

Theorem 4.2. *Under the same assumption in Theorem 4.1, let*

$$X_m \sim \nu_n,$$

where ν_n is defined in (13). Fix $k > 2$ and $r > 1$. Let $\tilde{L}_n = \max\{N_n, |\mathcal{N}_n|\}$, where $\mathcal{N}_n \subset D$ is a $R(x)^{-1}$-weighted $(kR_\mu(V_n))^{-1}$-covering of D. If $M_n \geq (r+1)G_{\mathcal{N}_n}\tilde{L}_n \log \tilde{L}_n$, then with probability at least $1 - N_n^{-r}$,

$$\|v\|_\infty \leq \frac{k}{k-2}\|v\|_{\mathcal{A}_n,\infty}.$$

Before giving the proof, we note that Theorem 4.2 combined with assumption (15) immediately implies the following result:

Theorem 4.3. *Under the same condition as in Theorem 4.2 and (15), with probability 1, the sequence $\{\mathcal{A}_n\}$ generated by ν_n forms an AM with $a = p^* + \tau$ and $b = 0$ for any $\tau > 0$.*

Proof. Assumption (15) implies that for any $\tau > 0$, there exists a sequence $\{\mathcal{N}_n\}$ such that $\mathcal{N}_n$ is an f-weighted $(3R_\mu(V_n))^{-1}$-covering of D and

$$\limsup_{n\to\infty} \frac{\log\left(|\mathcal{N}_n| G_{\mathcal{N}_n}\right)}{\log N_n} < p^* + \frac{\tau}{2}.$$

The corresponding M_n derived in Theorem 4.2 satisfies

$$\begin{aligned}\limsup_{n\to\infty} \frac{M_n}{\log N_n} &= \limsup_{n\to\infty} \left(\frac{\log(r+1)}{\log N_n} + \frac{\log\left(|\mathcal{N}_n| G_{\mathcal{N}_n}\right)}{\log N_n} + \frac{\log\log\left(|\mathcal{N}_n| + N_n\right)}{\log N_n} \right) \\ &< p^* + \frac{\tau}{2}.\end{aligned}$$

Thus, there exists some constant C_τ such that $M_n \le C_\tau N_n^{p^*+\tau}$. On the other hand, Theorem 4.2 tells us that for every n with sampling size M_n,

$$\Pr\left[\|v\|_\infty > 3\|v\|_{\mathcal{A},\infty}\right] \le N_n^{-r} \le n^{-r}.$$

An application of the Borel-Cantelli lemma finishes the proof. □

Proof of Theorem 4.2. Assume that $\|v\|_{2,\mu} = \sum_{i\in[N_n]} \langle v, v_i\rangle_\mu^2 = 1$. Let us consider an f-weighted η-covering of D with

$$f = R_n(x)^{-1}, \tag{32}$$

where η is defined in (28). Note that every η-covering of D is also an f-weighted η-covering of D, see Section 2.4. With some abuse of notation, from now on let $\mathcal{N}_n$ be an f-weighted η-covering of D. A similar computation as before shows that if $\mathcal{A}_n \subset D$ has non-empty intersection with $B_{r(y,\eta)}(y)$ for every $y \in \mathcal{N}_n$, then

$$\|v\|_\infty \le \frac{k}{k-2} \|v\|_{\mathcal{A}_n,\infty}. \tag{33}$$

To see this, note that the non-empty intersection property implies that for $x \in \operatorname{argmax}_{u\in D} v(u)$, there exist $y \in \mathcal{N}_n$ and $x' \in \mathcal{A}_n$ such that $\max\{|y-x|, |y-x'|\} \le r(y,\eta)$. Therefore,

$$\begin{aligned}\|v\|_{\mathcal{A}_n,\infty} \ge v(x') &\ge v(x) - |v(x) - v(y)| - |v(y) - v(x')| \\ &\overset{(31)}{\ge} \|v\|_\infty - 2r(y,\eta) \max_{z\in B_{r(y,\eta)}(y)} R_n(z) \\ &\overset{(12)}{=} \|v\|_\infty - 2\eta F_{r(y,\eta)}(y) \max_{z\in B_{r(y,\eta)}(y)} R_n(z) \\ &= \|v\|_\infty - \frac{2}{k} \\ &\overset{\|v\|_\infty \ge \|v\|_{2,\mu}=1}{>} \|v\|_\infty - \frac{2\|v\|_\infty}{k},\end{aligned}$$

which shows (33).

We next construct such sets $\mathcal{A}_n$ satisfying the non-empty intersection property using random sampling. To this end, we draw X_m independently from a probability distribution ν_n which is defined in (13). It is easy to check that for $y \in \mathcal{N}_n$,

$$\Pr\left[X_m \in B_{r(y,\eta)}(y)\right] = \frac{\int_{B_{r(y,\eta)}(y)} R_n^d(x)dx}{\int_D R_n^d(x)dx} \ge \frac{\int_{B_{r(y,\eta)}(y)} R_n^d(x)dx}{\sum_{z\in\mathcal{N}_n} \int_{B_{r(z,\eta)}(z)} R_n^d(x)dx}. \tag{34}$$

The numerator of the last term in (34) can be bounded from below as

$$\begin{aligned}\int_{B_{r(y,\eta)}(y)} R_n^d(x)dx &\geq |B_{r(y,\eta)}(y)| \min_{x\in B_{r(y,\eta)}(y)} R_n^d(x)\\ &= \sigma_d (r(y,\eta))^d \min_{x\in B_{r(y,\eta)}(y)} R_n^d(x)\\ &\overset{(12)}{=} \sigma_d \eta^d F_{r(y,\eta)}(y)^d \min_{x\in B_{r(y,\eta)}(y)} R_n^d(x)\\ &= \sigma_d \left(\frac{\min_{x\in B_{r(y,\eta)}(y)} R_n(x)}{k \max_{x\in B_{r(y,\eta)}(y)} R_n(x)}\right)^d \geq \sigma_d \frac{1}{k^d G_{\mathcal{N}_n}},\end{aligned} \tag{35}$$

where σ_d is the volume of the unit ball in $\mathbb{R}^d$. The denominator of the last term in (34), on the other hand, can be bounded from above by

$$\begin{aligned}\sum_{z\in\mathcal{N}_n} \int_{B_{r(z,\eta)}(z)} R_n^d(x)dx &\leq \sum_{z\in\mathcal{N}_n} |B_{r(z,\eta)}(z)| \max_{x\in B_{r(z,\eta)}(z)} R_n^d(x)\\ &= \sum_{z\in\mathcal{N}_n} \sigma_d \left(r(z,\eta) \max_{x\in B_{r(z,\eta)}(z)} R_n(x)\right)^d \overset{(12)}{=} \frac{|\mathcal{N}_n|\sigma_d}{k^d}.\end{aligned} \tag{36}$$

Plugging (35) and (36) into (34) yields

$$\Pr\left[X_m \in B_{r(y,\eta)}(y)\right] \geq \frac{1}{G_{\mathcal{N}_n}|\mathcal{N}_n|}.$$

By a similar reasoning one can show that by taking $M_n = (r+1)G_{\mathcal{N}_n}|\mathcal{N}_n|\log|\mathcal{N}_n|$, with probability at least $1-|\mathcal{N}_n|^{-r}$, $\{X_m\}$ intersects every ball in the covering of $\mathcal{N}_n$, completing the proof. □

5 Numerical simulations

We provide simulations to demonstrate the effectiveness of random sampling strategies for generating WAMs and AMs proposed in the previous sections. Particularly, we will compare the performance of the R_n-weighted sampling with the uniform sampling in a specific case in 1D through an examination of the f-weighted covering number.

Take $D = [-1,1]$ and V_n as the the space of Jacobi polynomials with parameters (α,β) and degree less than or equal to n, i.e.,

$$\frac{d\mu}{dx} \propto (1-x)^\alpha (1+x)^\beta \qquad\qquad x \in (-1,1).$$

In the rest of the discussion we assume $\alpha = \beta = \gamma > -1$, so that V_n becomes a single-index family of orthogonal polynomials on $[-1,1]$. For example, taking $\gamma = 0, \mp 0.5$ yields the Legendre polynomials and the first/second type of Chebyshev polynomials, respectively.

In the first simulation, we choose $n = 12$ and $\gamma = -0.5, 0, 0.5$ and 1. The meshes are generated independently by uniform sampling on $[-1,1]$, the $R_n(x)$-weighted sampling $\nu_n \propto R_n(x)$, μ and μ_V, respectively. For a fixed mesh, we generate 5000 random unit L^2_μ-norm functions $v \in V_n$; i.e., with $\{v_i\})_{i=1}^n$ an orthonormal basis for V_n, we sample expansion coefficients uniformly from the unit sphere in $\mathbb{R}^n$, i.e.,

$$(\langle v, v_i\rangle)_{i\in[n]} \sim \mathsf{Unif}(\mathbb{S}^{n-1}).$$

We record the maximal ratio between $\|v\|_\infty$ and $\|v\|_{\mathcal{A}_n,\infty}$ among the 5000 functions for each mesh size and treat it as an approximate estimate for C_n defined in (1). The experiment is repeated 100 times with its 0.05-0.5-0.95 quantiles reported in Figure 1.

Overall, the R_n-weighted sampling exhibits decent performance, but slightly underperforms compared to sampling with μ and μ_V when $\gamma = -0.5$. This is perhaps not surprising since in

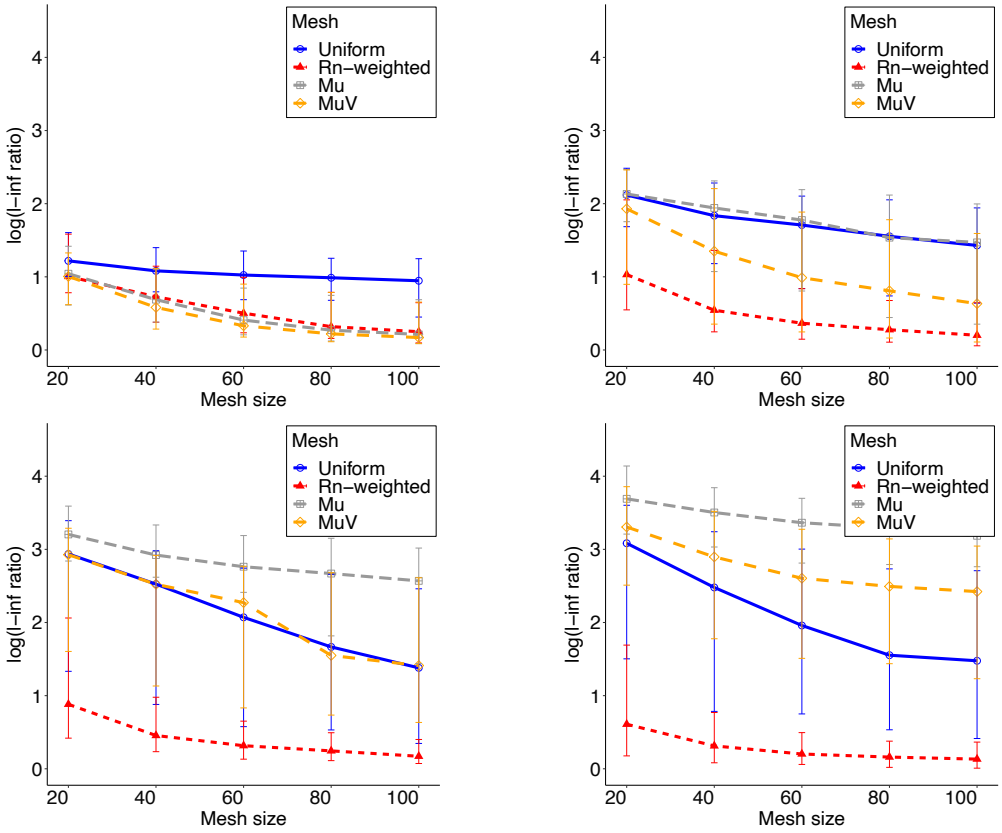

Figure 1: Empirically computed $\sup_{v\in V_n} \frac{\|v\|_\infty}{\|v\|_{\mathcal{A}_n,\infty}}$ (log-scale) on meshes generated by the uniform sampling, the R_n-weighted sampling, μ and μ_V. $\gamma = -0.5$ (Top Left), $\gamma = 0$ (Top Right), $\gamma = 0.5$ (Bottom Left) and $\gamma = 1$ (Bottom Right).

this case μ is the arcsine measure, which is the asymptotic limit of optimal measures [1], and hence should exhibit excellent approximation properties. When $\gamma = 0.5$ and $\gamma = 1$, R_n-weighted sampling demonstrates substantial performance improvement over the alternative methods. To understand why this happens, we analyze the bound on M_n obtained in Theorem 4.1 and Theorem 4.2, which boils down to estimating covering numbers. In the sequel, we will derive a useful estimate for the f-weighted $(kR_\mu(V_n))^{-1}$-covering number for different n's and γ's and compare it to the classical covering number with the same parameters. Let us fix $k = 3$ for convenience.

Since $D = [-1, 1]$, its classical $(kR_\mu(V_n))^{-1}$-covering number can be explicitly computed as $\lceil kR_\mu(V_n)\rceil$. To estimate the f-weighted covering number (where f is the same as defined in (32)), consider an interval partition $\{I_j\}_{j\in[J]}$ of the domain: $D = [-1, 1] = \cup_{j\in[J]} I_j$. For $y \in I_j$,

$$\min\left\{\frac{\min_{z\in I_j} f(z)}{k}, \operatorname{dist}(y, \partial I_j)\right\} \le r(y, (kR_\mu(V_n))^{-1}) \le \frac{\max_{z\in I_j} f(z)}{k},$$

where ∂I_j is the set of the boundary points of I_j. It is easy to check that if

$$\frac{2\max_{z\in I_j} f(z)}{k} = 2\left(k\min_{z\in I_j} R_n(z)\right)^{-1} \le |I_j| \qquad \forall j \in [J],$$

then the f-weighted $(kR_\mu(V_n))^{-1}$-covering number is bounded from above by

$$\sum_{j\in[J]} \left\lceil \frac{k|I_j|}{2\min_{x\in I_j} f(x)} \right\rceil, \tag{37}$$

and bounded from below by

$$\sum_{j\in[J]} \frac{k|I_j|}{2\max_{x\in I_j} f(x)}. \tag{38}$$

We now construct a partition $\{I_j\}_{j=1}^J$ of $[-1,1]$ via level sets of the weight f. Let $M_f = \max_{u\in D} f(u)$, $m_f = \min_{u\in D} f(u)$, and $J = \left\lceil \log_2 \frac{M_f}{m_f} \right\rceil$. Write

$$[-1,1] = \bigcup_{j\in[J]} \underbrace{\{x \in D : 2^{-j} M_f < f(x) \le 2^{-j+1} M_f\}}_{I_j} = \bigcup_{j\in[J]} \bigcup_{s\in[r_j]} I_{j,s}, \tag{39}$$

where $I_{j,s}$ are disjoint intervals such that $I_j = \cup_{s\in[r_j]} I_{j,s}$, and r_j is some finite integer. This is possible since $f(x)$ is a positive smooth function on $[-1,1]$, see Figure 2. We have partitioned the interval $[-1,1]$ by $I_{j,s}$; note that each I_j is not an interval, but a union of several disjoint intervals. r_j denotes the number of connected components in I_j. Thanks to the dyadic choice of level sets, (37) and (38) only differ by a multiple constant 2 (ignoring the ceiling effect), suggesting that (37) is a tight estimate. With such a construction, (37) can be further bounded as

$$\sum_{j\in[J]} \sum_{s\in[r_j]} \left\lceil \frac{k|I_{j,s}|}{2\max\{2^{-j}M_f, m_f\}} \right\rceil \le \sum_{j\in[J]} \left(\left\lceil \frac{k|I_j|}{2\max\{2^{-j}M_f, m_f\}} \right\rceil + r_j \right). \tag{40}$$

The extra r_j term above comes from taking into account the ceiling effect as well as boundary terms. Although the dyadic choice of I_j in (39) may not be the best $(kR_\mu(V_n))^{-1}$-covering, it guarantees that $G_{\mathcal{N}_n} \le 2$, which together with (40) yields an upper bound for the M_n in Theorem 4.2. This allows us to numerically obtain an upper bound for the exponent a which is defined in Definition 2.1 but not specified in Theorem 4.2.

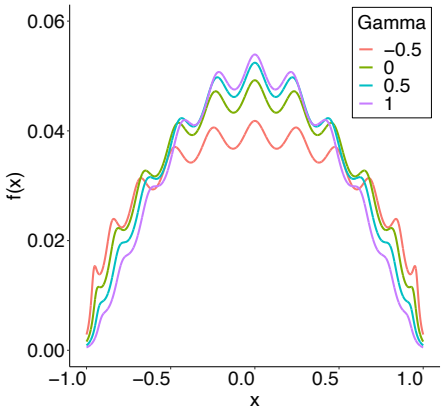

Figure 2: Graph of the weight function $f(x) = R_n(x)^{-1}$ when $V_n = P_{12}$ and $\gamma = -0.5, 0, 0.5, 1$. In all cases, the minimum of $f(x)$ is positive and achieved at the boundary of $[-1,1]$.

We now use the idea above to estimate the required mesh size M_n given in Theorem 4.1 ($M_n^{(\text{Uniform})} = 3L_n \log L_n$) and Theorem 4.2 ($M_n^{(R_n\text{-weighted})} = 3G_{\mathcal{N}_n} \tilde{L}_n \log \tilde{L}_n$) for the uniform and R_n-weighted sampling. Since f depends on both n and γ, we will vary both parameters in the simulation to investigate their corresponding effect on M_n. Particularly, we will take $\gamma \in \{-1 + i/2 : i \in [12]\}$ and $n \in \{20s : s \in [15]\}$. Note that M_n quickly becomes extremely large as n increases. Thus we compute $\log M_n$ instead of M_n for each n. Moreover, the exponent a, if it exists, is close to the slope of the scatterplot of $(\log n, \log M_n)$, which can be estimated via least-squares line fit. In Figure 3, we plot $\log M_n$ against $\log n$ for a selection of γ's. The estimated slopes a for all cases of γ can be found in Table 2.

To compare the efficiency between the uniform sampling and the R_n-weighted one, note that for a fixed error (n^{-2} for instance), the ratio between the required sampling size M_n in Theorem 4.1 and Theorem 4.2 is

$$\frac{M_n^{(\text{Uniform})}}{M_n^{(R_n\text{-weighted})}} \approx \frac{n^{a^{(\text{Uniform})}}}{n^{a^{(R_n\text{-weighted})}}} = n^{a^{(\text{Uniform})} - a^{(R_n\text{-weighted})}},$$

where $a^{(\cdot)}$ is the exponent defined in Definition 2.1 for the given method. This motivates us to define the relative efficiency of the R_n-weighted sampling over the uniform sampling as the difference between these exponents,

$$\text{relative efficiency} := a^{(\text{Uniform})} - a^{(R_n\text{-weighted})},$$

which can be approximated using Table 2. We plot the relative efficiency of the R_n-weighted sampling for different γ's in Figure 3. We see that R_n-weighted sampling is polynomially better than uniform sampling in terms of the required sampling size. Such effect becomes more and more prominent when γ increases from -0.5 to 1, which is consistent with Figure 1.

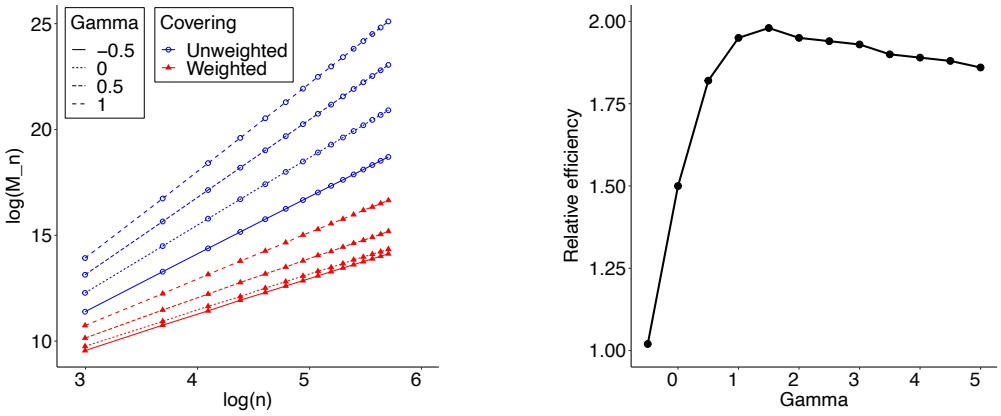

Figure 3: Left: Scatterplot of $(\log n, \log M_n)$ for both the uniform sampling (Unweighted) and the R_n-weighted (Weighted) sampling in the cases $\gamma = -0.5, 0, 0.5, 1$. Right: Relative efficiency of the R_n-weighted sampling over the uniform sampling.

Weight parameter γ	**Uniform sampling** Estimated exponent a	R_n**-weighted sampling** Estimated exponent a
-0.5	2.70	1.68
0	3.19	1.69
0.5	3.67	1.85
1.0	4.14	2.19
1.5	4.61	2.63
2.0	5.07	3.12
2.5	5.52	3.58
3.0	5.97	4.04
3.5	6.41	4.51
4.0	6.85	4.96
4.5	7.28	5.40
5.0	7.71	5.85

Table 2: Least-squares estimate for a for both uniform sampling (Unweighted) and R_n-weighted (Weighted) sampling in the case of different γ's.

Although we have observed that R_n-weighted sampling is theoretically and empirically superior to uniform sampling, our analysis suggests that this improvement has limits. For example,

note that $G_{\mathcal{N}_n} \geq 1$ and by definition, the f-weighted $(kR_\mu(V_n))^{-1}$-covering number is bounded from below (ignoring the ceiling effect) by

$$\tilde{L}_n \geq \frac{k}{M_f} = \frac{m_f}{M_f} L_n.$$

Therefore, the relative efficiency of the R_n-weighted sampling over the uniform sampling is bounded from above by

$$\begin{aligned} \text{relative efficiency} &= \frac{\log M_n^{(\text{Uniform})}}{\log n} - \frac{\log M_n^{(R_n\text{-weighted})}}{\log n} \\ &= \frac{\log(3L_n \log L_n)}{\log n} - \frac{\log(3G_{\mathcal{N}_n}\tilde{L}_n \log \tilde{L}_n)}{\log n} \\ &\lesssim \frac{\log\left(M_f/m_f\right)}{\log n}. \end{aligned} \tag{41}$$

If $M_f/m_f = \mathcal{O}(n^\beta)$ for some $\beta > 0$, then the above bound is asymptotically equivalent to

$$\lim_{n\to\infty} \frac{\log(M_f/m_f)}{\log n} = \beta.$$

In this case, the R_n-weighted sampling is at most n^β more efficient than uniform sampling based on the covering number analysis.

Acknowledgements

A. Narayan thanks Norm Levenberg and Sione Ma'u for a careful reading of an early draft, and for providing several comments that greatly improved the quality of the manuscript. Y. Xu and A. Narayan are partially supported by NSF DMS-1848508. This material is based partially upon work supported by the National Science Foundation under Grant No. DMS-1439786 and by the Simons Foundation Grant No. 50736 while A. Narayan was in residence at the Institute for Computational and Experimental Research in Mathematics in Providence, RI, during the "Model and dimension reduction in uncertain and dynamic systems" program.

References

[1] T. Bloom, L. Bos, N. Levenberg, and S. Waldron. "On the Convergence of Optimal Measures". *Constructive Approximation* 32.1 (2010), pp. 159–179. ISSN: 0176-4276. DOI: 10.1007/s00365-009-9078-7.

[2] T. Bloom, L. P. Bos, J.-P. Calvi, and N. Levenberg. "Polynomial interpolation and approximation in C^d". *Annales Polonici Mathematici* 106 (2012), pp. 53–81. ISSN: 0066-2216, 1730-6272. DOI: 10.4064/ap106-0-5.

[3] L. Bos, J.-P. Calvi, N. Levenberg, A. Sommariva, and M. Vianello. "Geometric weakly admissible meshes, discrete least squares approximations and approximate Fekete points". *Mathematics of Computation* 80.275 (2011), pp. 1623–1638. ISSN: 0025-5718.

[4] J.-P. Calvi and N. Levenberg. "Uniform approximation by discrete least squares polynomials". *Journal of Approximation Theory* 152.1 (2008), pp. 82–100. ISSN: 0021-9045. DOI: 10.1016/j.jat.2007.05.005.

[5] A. Chkifa, A. Cohen, G. Migliorati, F. Nobile, and R. Tempone. "Discrete least squares polynomial approximation with random evaluations - application to parametric and stochastic elliptic PDEs". *ESAIM: Mathematical Modelling and Numerical Analysis* 49.3 (2015), p. 23. DOI: 10.1051/m2an/2014050.

[6] A. Cohen, M. A. Davenport, and D. Leviatan. "On the Stability and Accuracy of Least Squares Approximations". en. *Foundations of Computational Mathematics* 13.5 (2013), pp. 819–834. ISSN: 1615-3375, 1615-3383. DOI: 10.1007/s10208-013-9142-3.

[7] A. Cohen and G. Migliorati. "Optimal weighted least-squares methods". *SMAI Journal of Computational Mathematics* 3 (2017). arxiv:1608.00512 [math.NA], pp. 181–203. ISSN: 2426-8399. DOI: 10.5802/smai-jcm.24.

[8] J. Dick and F. Pillichshammer. *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. English. 1 edition. New York, NY: Cambridge University Press, 2010. ISBN: 978-0-521-19159-3.

[9] W. Gautschi and S. Li. "A set of orthogonal polynomials induced by a given orthogonal polynomial". en. *aequationes mathematicae* 46.1-2 (1993), pp. 174–198. ISSN: 0001-9054, 1420-8903. DOI: 10.1007/BF01834006.

[10] J. Hampton and A. Doostan. "Coherence motivated sampling and convergence analysis of least squares polynomial Chaos regression". *Computer Methods in Applied Mechanics and Engineering* 290 (2015), pp. 73–97. DOI: 10.1016/j.cma.2015.02.006.

[11] S. Hussung. "Pluripotential Theory Associated with Convex Bodies and Related Numerics". en. PhD thesis. Indiana University, 2020.

[12] G. Migliorati. "Multivariate Markov-type and Nikolskii-type inequalities for polynomials associated with downward closed multi-index sets". *Journal of Approximation Theory* 189 (2015), pp. 137–159. DOI: 10.1016/j.jat.2014.10.010.

[13] A. Narayan. "Computation of induced orthogonal polynomial distributions". *Electronic Transactions on Numerical Analysis* 50 (2018). arXiv:1704.08465 [math], pp. 71–97. DOI: 10.1553/etna_vol50s71.

[14] A. Narayan, J. Jakeman, and T. Zhou. "A Christoffel function weighted least squares algorithm for collocation approximations". *Mathematics of Computation* 86.306 (2017). arXiv: 1412.4305 [math.NA], pp. 1913–1947. ISSN: 0025-5718, 1088-6842. DOI: 10.1090/mcom/3192.