Near-optimal sampling strategies for multivariate function approximation on general domains

Ben Adcock and Juan M. Cardenas
Department of Mathematics
Simon Fraser University
Canada
March 18, 2022

Abstract

In this paper, we address the problem of approximating a multivariate function defined on a general domain in $d$ dimensions from sample points. We consider weighted least-squares approximation in an arbitrary finite-dimensional space $P$ from independent random samples taken according to a suitable measure. In general, least-squares approximations can be inaccurate and ill-conditioned when the number of sample points $M$ is close to $N = \dim(P)$. To counteract this, we introduce a novel method for sampling in general domains which leads to provably accurate and well-conditioned approximations. The resulting sampling measure is discrete, and therefore straightforward to sample from. Our main result shows near-optimal sample complexity for this procedure; specifically, $M = \mathcal{O}(N \log(N))$ samples suffice for a well-conditioned and accurate approximation. Numerical experiments on polynomial approximation in general domains confirm the benefits of this method over standard sampling.

1 Introduction

In this paper, we consider the problem of approximating a multivariate function $f : \Omega \to \mathbb{C}$ of $d \geq 1$ variables whose domain $\Omega \subseteq \mathbb{R}^d$ may be irregular. This problem arises in many applications in computational science and engineering, and presents two main challenges. First, the well known curse of dimensionality, and second, the potential irregularity of the domain $\Omega$. While there has been significant progress made towards mitigating the former (see §1.1), the majority of this work has focused on the case of tensor-product domains, for instance, the unit hypercube $\Omega = [-1,1]^d$. Far less attention has been paid to the case of irregular domains.

Recently, in [4] the first author developed a framework for polynomial approximation of smooth functions in general domains in $d$ dimensions. The approach, known as polynomial frame approximation, is based on regularized least-squares approximation using orthonormal polynomials on a bounding hypercube and random sampling from the restriction of the orthogonality measure to $\Omega$.

For certain domains and polynomial spaces, this procedure has provable bounds on the sample complexity; that is, the scaling between the dimension of the approximation space $N$ and the number of pointwise samples $M$ which is sufficient to guarantee a well conditioned and accurate approximation. While these bounds are independent of the dimension $d$, and therefore ameliorate the curse of dimensionality in the number of function evaluations, the best known bounds are quadratic in $N$, i.e. $M = \mathcal{O}(N^2 \log(N))$, and are known to hold only for domains possessing the so-called $\lambda$-rectangle property and polynomial spaces corresponding to lower sets of multi-indices.
The reason for this can be traced to the choice of measure from which the sample points are drawn. In the case of Legendre polynomials, for instance, this is simply the uniform measure over Ω, which is known to be a relatively poor distribution for polynomial approximation.

In recent works [8, 11, 19], it has been shown how construct a sampling measure depending on the space P which leads to the near-optimal scaling $M = O(N \log(N))$, where $N$ is the dimension of $P$. Note that $P$ can be an arbitrary finite-dimensional subspace in this setup; it need not be a space of polynomials. Unfortunately, the practical implementation of this approach requires two ingredients: first, an orthonormal basis for $P$, and second, a tensorial structure for the corresponding basis functions. The latter is used in order to efficiently sample from the constructed measure. Orthonormal polynomials on hypercubes typically exhibit both these qualities; for instance, the Legendre polynomials on $[-1, 1]^d$ are simply the tensor-products of the univariate Legendre polynomials on $[-1, 1]$, thus both tensorial and easy to construct. However, for irregular domains, neither property holds in general.

In this work, we combine the ideas of [4] and [11], as well as those of [19], to construct a weighted least-squares approximation on general domains with the near-optimal sample complexity $M = O(N \log(N))$. Note that our main results guarantee, under this scaling, an error bound relating the $L^2$-norm error to a best approximation term in a certain weighted sup-norm (this is not quite optimal – see Remark 3.2). Our method is based on three steps. First, using the results of [4] we generate a fine grid of $K \gg N$ points over the domain $\Omega$. Throughout, we assume this step is computationally feasible; see §6 for further discussion on this topic. Second, starting from a nonorthogonal basis for the approximation space – for example, as in [4, 6], the restriction of an orthonormal basis on a bounding box to $\Omega$ – we construct an orthonormal basis with respect to the corresponding discrete measure supported on the grid (note that the approach in [4, 6] does not seek to orthogonalize the original basis). Third, we use the ideas of [11, 19] to generate a near-optimal sampling measure. Unlike in these works, the resulting sampling measure is discrete, supported on the grid of $K$ points, and therefore straightforward to sample randomly from. Following ideas from [19], we present two versions of our approach. The first method (Method 1) considers a fixed approximation space $P$, while the second (Method 2) considers a sequence of nested spaces $P_1 \subset P_2 \subset P_3 \subset \ldots$. The second method has the benefit of being adaptive: all the sample points used to compute the approximation in the space $P_i$ are recycled when computing the approximation in $P_{i+1}$.

To demonstrate the effectiveness of these two methods, we present numerical experiments showing polynomial approximation on general domains in arbitrary dimensions. For many domains in various dimensions, the new sampling methods achieve better accuracy and stability than when drawing samples from the uniform measure, as was done in [4]. Furthermore, the adaptive method (Method 2) leads to no deterioration in accuracy or stability over Method 1.

1.1 Related work

Motivated by applications in uncertainty quantification and parametric PDEs, least-squares polynomial approximation of high-dimensional, smooth functions has received significant attention over the last ten years. The majority of works have focused on tensor-product domains. Besides [4], mentioned above, very few works have considered the question of general domains $\Omega$. Note that [4], based on ideas of [6, 5], uses Singular Value Decomposition (SVD) followed by thresholding of the singular values to address the nonorthogonality of the approximation system. Conversely, in this paper, we use QR decomposition to explicitly orthogonalize the basis. The resulting orthonormal basis is used to construct the near-optimal sampling measure and then to compute the final approximation.
Since it is often a critical constraint in practice, a major focus of previous work on least-squares polynomial approximation has been quantifying the sample complexity when the samples are drawn randomly from the orthogonality measure of the polynomial basis employed. See [10, 17, 18, 21] and references therein. Unfortunately, these methods tend to have superlinear sample complexity. Approaches at designing sampling measures which lead to log-linear sample complexity have been considered in, for instance, [22], the aforementioned works [11, 19] and [8].

We remark in passing that least-squares approximation, as we consider in this paper, is but one approach for polynomial approximation in high dimensions. A related, but distinct, line of work uses compressed sensing techniques for this problem. See [1, 2, 3, 9, 13, 14, 23] and references therein. This approach is quite powerful, since, unlike least squares, it does not require one to specify a priori the approximation space $P$. However, it is not yet known how to perform (provably) optimal sampling in the compressed sensing setting.

Finally, we note that a similar approach to Method 1 of this paper, with corresponding theoretical analysis, has also been developed simultaneously by Migliorati [20].

1.2 Outline

The outline of the remainder of this paper is as follows. We first summarize our two methods, Method 1 and Method 2, in §2. In §3 we present the main theoretical analysis of these methods. Proofs of the results presented in this section are given in §4. We conclude in §5 with numerical examples.

2 Summary of the methods

We now present our two main methods: Nonadaptive sampling (Method 1) and Adaptive sampling (Method 2).

2.1 Method 1

In Method 1 (shown in the box below) we compute an approximation $\tilde{f} \in P$ to a function $f$ from a fixed subspace $P$ using the set of samples $\{f(y_i)\}_{i=1}^M$. The sample points $y_1, \ldots, y_M$ are drawn independently and identically from the grid $Z$ according to probability distribution $\pi$. In other words, $y_i \sim \mu$, where $\mu$ is the discrete sampling measure

$$d\mu(y) = \sum_{i=1}^K \pi_i \delta(y - z_i), \quad y \in \Omega.$$ 

Observe that this measure is precisely

$$d\mu(y) = \sum_{i=1}^K \left( \frac{1}{N} \sum_{i=1}^N |\phi_i(y)|^2 \right)^{-1} \delta(y - z_i).$$

The function $\sum_{i=1}^N |\phi_i(y)|^2$ is the reciprocal of the Christoffel function of $P$ [24], which was previously identified in [11] as a suitable measure from which to obtain optimal sampling.

Note that the reduced QR decomposition in Step 3 of Method 1 (and likewise for Method 2) refers to the decomposition $B = QR$, where $Q$ is $K \times N$ with orthonormal columns, i.e. $Q^*Q = I$ and $R$ is $N \times N$ and upper triangular.
Method 1. Nonadaptive sampling for general domains

**Inputs:** Domain \( \Omega \), probability measure \( \rho \) over \( \Omega \), function \( f \in L^\infty(\Omega) \).
Finite-dimensional subspace \( P \subset L^2(\Omega, \rho) \) of dimension \( N \).
Basis \( \{ \psi_1, \ldots, \psi_N \} \) for \( P \) (not necessarily orthogonal).
Number of sample points \( M \geq N \), fine grid size \( K \geq N \).

**Step 1:** Draw \( K \) points \( Z = \{ z_i \}_{i=1}^K \) independently from \( \rho \).

**Step 2:** Construct the \( K \times N \) matrix \( B = \{ \psi_j(z_i)/\sqrt{K} \}_{i,j=1}^{K,N} \) and check whether or not \( B \) has full rank (\( \text{rank}(B) = N \)). If not, go back to Step 1. Else proceed to Step 3.

**Step 3:** Compute the reduced QR decomposition \( B = QR \), where \( Q = \{ q_{ij} \}_{i,j=1}^{K,N} \in \mathbb{C}^{K \times N} \) and \( R \in \mathbb{C}^{N \times N} \). Define the probability distribution \( \pi = \{ \pi_i \}_{i=1}^K \) on \( \{ 1, \ldots, K \} \) by

\[
\pi_i = \frac{1}{N} \sum_{j=1}^N |q_{ij}|^2, \quad i = 1, \ldots, K.
\]

**Step 4:** Draw \( M \) integers \( i_1, \ldots, i_M \) independently from \( \pi \), define

\[
A = \left\{ \frac{q_{i_j,k}}{\sqrt{M \pi_{i_j}}} \right\}_{j,k=1}^{M,N} \in \mathbb{C}^{M \times N}, \quad b = \left\{ \frac{f(z_{i_j})}{\sqrt{MK \pi_{i_j}}} \right\}_{j=1}^M \in \mathbb{C}^M,
\]

and compute \( c = \text{argmin}_{x \in \mathbb{C}^N} \| Ax - b \|_2 \).

**Output:** The approximation \( \tilde{f}(y) = \sum_{i=1}^N c_i \phi_i(y) \), where \( \phi_i(y) = \sum_{j=1}^i (R^{-*})_{ij} \psi_j(y) \).

Several remarks are in order. First, both Method 1 and Method 2 (described next) assume it is possible to draw samples from the probability measure \( \rho \). This can be achieved via, for example, rejection sampling, as we do in our experiments later. However, this may not be feasible in all settings, depending on the problem at hand. Second, in \( \S 3.6 \) we derive guarantees on \( K \) which ensure \( B \) has full rank with high probability (see Step 2). Naturally, in practice, if \( B \) fails to be full rank then, rather than throwing away the current points, one may instead prefer to increase \( K \) and draw additional points until full rankness is achieved. See [20] for further discussion on this issue. Third, note that if \( \tilde{f} \) is only sought on the fine grid, then the computation of the functions \( \phi_i \) in the final stage is unnecessary. Since

\[
\{ \tilde{f}(z_i) \}_{i=1}^K = \sqrt{K} Q c,
\]
evaluating \( \tilde{f} \) on this grid involves only a simple matrix-vector multiplication. Fourth, we remark in passing that the scalings of the rows of \( A \) and \( b \) are to ensure good conditioning of \( A \), under suitable conditions on \( M \) and \( N \). See Theorem 2.1 below. Finally, we note the computational cost is (assuming the \( \psi_j \)'s are cheap to evaluate) dominated by the cost of computing the QR decomposition of \( B \), which involves an offline cost of \( O(KN^2) \) flops, and solving the least-squares problem to obtain \( c \), which involves an online cost of \( O(MN^2) \) flops.

Two main questions we investigate in this paper are how large \( M \) needs to be in comparison to \( N \) and how large \( K \) needs to be in comparison to \( N \). Note that the former pertains to the sample complexity of the method. As we show later, with the probability distribution defined in Step 3, the log-linear scaling \( M \asymp N \log(N) \) is sufficient for a well conditioned approximation
which also accurately approximates \( f \) over the fine grid \( Z \). Besides better sample complexity, this also compares favourably with the method of [3] in terms of the online computational cost, which is \( \mathcal{O}(N^3 \log(N)) \) for any domain, whereas the cost in [3] is \( \mathcal{O}(N^4 \log(N)) \) and only for certain domains (i.e. those for which the sample complexity is provably quadratic in \( N \)).

To ensure the approximation is also accurate over \( \Omega \), we need \( K \) to be sufficiently large in relation to \( N \). Currently, we have no complete answer for general domains and spaces \( P \). However, if \( \rho \) is the uniform measure and \( P \) is a polynomial subspace based on a so-called lower set of multi-indices (as is typical in practice), then we show that \( K \approx N^2 \lambda^{-1} \log(N) \) is sufficient, provided the domain \( \Omega \) has the so-called \( \lambda \)-rectangle property. We discuss this further in \S 3.6.

### 2.2 Method 2

Method 1 has the limitation that if the subspace \( P \) is augmented to a larger space \( \tilde{P} \supset P \), the existing sample points \( y_1, \ldots, y_M \) are not sampled from the appropriate distribution for \( \tilde{P} \). In Method 2, following ideas of [19], we consider an adaptive procedure in which all samples are recycled as the \( P \) is increased.

In this method, the sample complexity for subspace \( P_t \) is \( M_t = k_t N_t \). As we show later, a suitable choice of \( k_t \) to ensure a sequence of well conditioned and accurate approximations is \( k_t \approx \log(N_t) \).

We note also that the QR decomposition computed in Step 3 need not be done from scratch at each step. One can use standard methods to update the decomposition according to the new columns added at each step. See, for example, [16, Chpt. 24].

### 2.3 Main theoretical results

Having presented our two methods, we now summarize their stability and accuracy, and in particular, the conditions on \( M \) and \( K \). This is the topic of the following two theorems. To this end, we now define the Nikolskii constant \( \mathcal{N}(P, \rho) \) as the smallest possible constant such that

\[
\|p\|_{L^\infty(\Omega)} \leq \mathcal{N}(P, \rho) \|p\|_{L^2(\Omega, \rho)}, \quad \forall p \in P.
\]  

(2.1)

See \[ 3.6 \] for further information.

**Theorem 2.1.** Consider the setup of Method 1. Suppose that \( \gamma, \delta \in (0, 1) \) and

\[
M \geq N \log(4N/\gamma) \left((1 + \delta) \log(1 + \delta) - \delta\right)^{-1},
\]

\[
K \geq (\mathcal{N}(P, \rho))^2 \log(2N/\gamma) \left((1 - \delta) \log(1 - \delta) + \delta\right)^{-1},
\]

where \( \mathcal{N}(P, \rho) \) is as in (2.1). Then the following holds with probability at least \( 1 - \gamma \):

(i) the matrix \( B \) is full rank,

(ii) the condition number of the matrix \( A \) satisfies \( \kappa(A) \leq \sqrt{\frac{1 + \delta}{1 - \delta}} \),

(iii) for any \( f \in L^\infty(\Omega) \) the approximation \( \tilde{f} \) is unique and satisfies

\[
\|f - \tilde{f}\|_{L^2(\Omega, \rho)} \leq \inf_{p \in P} \left\{ \|f - p\|_{L^2(\Omega, \rho)} + \frac{1}{1 - \delta} \|f - p\|_{Z, \pi} \right\},
\]

where \( \|g\|_{Z, \pi} = \max_{i=1,...,K} \left\{ \frac{\|g(z_i)\|}{\sqrt{K \pi_i}} \right\} \).
Theorem 2.2. Consider the setup of Method 2. Suppose that \( N_t \) is as in (2.1) with \( P = P_r \). Then the following holds with probability at least \( 1 - \gamma \).

For every \( t = 1, \ldots, r \),

(i) the matrix \( B \) of Step 3 is full rank,

(ii) the condition number of the matrix \( A \) in Step 6 satisfies \( \kappa(A) \leq \sqrt{\frac{1+\delta}{1-\delta}} \),
(iii) for any \( f \in L^\infty(\Omega) \) the approximation \( \tilde{f}^{(t)} \) is unique and satisfies

\[
\| f - \tilde{f}^{(t)} \|_{L^2(\Omega, \rho)} \leq \inf_{p \in P_t} \left\{ \| f - p \|_{L^2(\Omega, \rho)} + \frac{1}{1 - \delta} \| f - \tilde{p} \|_{Z, \pi, t} \right\},
\]

where \( \|g\|_{Z, \pi, t} = \max_{i=1, \ldots, K} \left\{ \left| g(z_i) \right| \sqrt{\frac{K}{M}} \sum_{l=1}^N \pi_l \right\} \).

3 Theoretical analysis

We now present our main analysis. To do so, we introduce a more general framework than that considered above, in which we consider three distinct quantities: an error measure \( \rho \), an orthogonality measure \( \tau \), and sampling measures \( \mu_1, \ldots, \mu_M \). Both Methods 1 and Methods 2 correspond to specific cases of this framework in which the orthogonality measure \( \tau \) is a discrete measure over the grid \( Z \). The difference between the two methods lies with the choices of the sampling measures \( \mu_i \). Note that this framework is general, and includes those of [11, 19] as special cases. Specifically, they correspond to the choice \( \tau = \rho \) and specific choices of the sampling measures \( \mu_i \) (see later).

The flexibility gained by allowing a distinct orthogonality measure \( \tau \) is what leads to Methods 1 and 2.

3.1 General setup

Consider the space \( L^2(\Omega, \rho) \) of square-integrable functions over a domain \( \Omega \subseteq \mathbb{R}^d \) with respect to a probability measure \( \rho \). We refer to \( \rho \) as the error measure: it gives the norm in which we measure the error of our approximation. Next, we define a second measure \( \tau \), the orthogonality measure, over \( \Omega \). We assume \( \tau \) is a probability measure, \( \int_\Omega d\tau = 1 \). This is the measure which we shall subsequently use to construct an orthonormal basis of the approximation space. Specifically, let \( P \subset L^\infty(\Omega) \) be the approximation space of dimension \( \dim(P) = N < \infty \). We write

\[
P = \text{span}\{\phi_1, \ldots, \phi_N\},
\]

where \( \{\phi_i\}_{i=1}^N \) is the corresponding orthonormal basis for \( P \) in \( L^2(\Omega, \tau) \). Finally, we define sampling measures \( \mu_1, \ldots, \mu_M \) over \( \text{supp}(\tau) \), the support of the measure \( \tau \). These are also probability measures. The \( i \)-th such measure \( \mu_i \) is the measure from which the \( i \)-th sample will be drawn. When later \( \tau \) is taken as a discrete measure, this means that the sampling measures will also be discrete measures. We also assume that there exists a function \( w \) that is positive and defined everywhere on \( \text{supp}(\tau) \) and satisfies

\[
\frac{1}{M} \sum_{i=1}^M d\mu_i(y) = \frac{1}{w(y)} d\tau(y), \quad \forall y \in \text{supp}(\tau).
\]

(3.1)

Note that this implies that \( \int_\Omega w^{-1} d\tau = 1 \).

We are now ready to define our approximation. Let \( M \geq N \) and draw \( M \) points \( y_1, \ldots, y_M \) independently, with \( y_i \) drawn according to the \( i \)-th sampling measure \( \mu_i \). We then define the weighted least-squares approximation of a function \( f \in L^\infty(\Omega) \) as

\[
\tilde{f} \in \arg\min_{p \in P} \left\{ \frac{1}{M} \sum_{i=1}^M w(y_i) | f(y_i) - p(y_i) |^2 \right\}.
\]

(3.2)

7
Write \( \tilde{f} = \sum_{i=1}^{N} c_i \phi_i \). Then this is equivalent to the algebraic weighted least-squares problem

\[
\mathbf{c} = (c_i)_{i=1}^{N} \in \arg\min_{\mathbf{x} \in \mathbb{C}^N} \| \mathbf{A} \mathbf{x} - \mathbf{b} \|_2,
\]

where

\[
\mathbf{A} = \left\{ \frac{1}{\sqrt{M}} \sqrt{w(y_i) \phi_j(y_i)} \right\}_{i,j=1}^{M,N} \in \mathbb{C}^{M \times N}, \quad \mathbf{b} = \left\{ \frac{1}{\sqrt{M}} \sqrt{w(y_i) f(y_i)} \right\}_{i=1}^{M} \in \mathbb{C}^M. \tag{3.3}
\]

For convenience, we now also define the discrete semi-inner product and semi-norm

\[
\langle f, g \rangle_{\Upsilon, w} = \frac{1}{M} \sum_{i=1}^{M} w(y_i) f(y_i) g(y_i), \quad \| f \|_{\Upsilon, w} = \sqrt{\langle f, f \rangle_{\Upsilon, w}}, \tag{3.4}
\]

where \( \Upsilon = \{ y_1, \ldots, y_M \} \). Note that \( \tilde{f} \) is can be expressed equivalently as

\[
\tilde{f} \in \arg\min_{p \in \mathcal{P}} \| f - p \|_{\Upsilon, w}. \tag{3.5}
\]

Given a domain \( \Omega \), an error measure \( \rho \) and an approximation space \( \mathcal{P} \), we are free to choose \( \tau \) and the \( \mu_i \). This raises the following question: how should one choose the orthogonality measure \( \tau \) and the sampling measures \( \mu_1, \ldots, \mu_M \)? There are two constraints to keep in mind. First, we wish to take as few samples \( M \) as possible. Second, we need probability measures \( \mu_i \) from which it is not computationally intensive to draw samples. We address the former determining how the error of the weighted least-squares approximation \( \tilde{f} \) depends on these quantities, and in particular, how the number of samples \( M \) influences the error bound. For the second, as noted, we construct \( \tau \) as a discrete measure over a suitable grid.

### 3.2 Error and sample complexity estimates

We first define the constant

\[
\mathcal{C} := \sup \left\{ \frac{\| p \|_{L^2(\Omega, \tau)}}{\| p \|_{\Upsilon, w}} : p \in \mathcal{P}, \ p|_{\text{supp}(\tau)} \neq 0 \right\}. \tag{3.6}
\]

Notice that \( \mathcal{C} < \infty \) if and only if \( \| \cdot \|_{\Upsilon, w} \) is a norm on \( \mathcal{P} \subset L^2(\Omega, \tau) \), which in turn is a necessary and sufficient condition for the least-squares problem (3.2) to have a unique solution.

This constant relates the orthogonality measure \( \tau \) to the sampling measures \( \mu_i \). We also need a constant relating the error measure \( \rho \) to \( \tau \). We define

\[
\mathcal{D} := \sup \left\{ \frac{\| p \|_{L^2(\Omega, \rho)}}{\| p \|_{L^2(\Omega, \tau)}} : p \in \mathcal{P}, \ p|_{\text{supp}(\rho)} \neq 0 \right\}. \tag{3.7}
\]

As above, notice that \( \mathcal{D} < \infty \) if and only if \( \| \cdot \|_{L^2(\Omega, \tau)} \) is a norm on \( \mathcal{P} \subset L^2(\Omega, \rho) \).

**Theorem 3.1.** Suppose that the constant \( \mathcal{C} \) defined in (3.6) satisfies \( \mathcal{C} < \infty \) and let \( f \in L^\infty(\Omega) \). Then the approximation \( \tilde{f} \) is unique and satisfies

\[
\| f - \tilde{f} \|_{L^2(\Omega, \tau)} \leq (1 + \mathcal{C}) \inf_{p \in \mathcal{P}} \| f - p \|_{\tau, w}.
\]

If in addition the constant \( \mathcal{D} \) defined in (3.7) satisfies \( \mathcal{D} < \infty \) then

\[
\| f - \tilde{f} \|_{L^2(\Omega, \rho)} \leq \inf_{p \in \mathcal{P}} \left\{ \| f - p \|_{L^2(\Omega, \rho)} + \mathcal{CD} \| f - p \|_{\tau, w} \right\},
\]

where \( \| g \|_{\tau, w} = \sup_{y \in \text{supp}(\tau)} \sqrt{w(y)} | g(y) | \).

8
See §4 for the proof. This result states that firstly the approximation error in $L^2(\Omega, \tau)$ is determined by the constant $C$, which relates the $L^2$-norm with respect to $\tau$ to the discrete $L^2$-norm over the sample points, and the best approximation error measured in the $\|\cdot\|_{\tau,w}$ norm, a weighted sup-norm over the support of $\tau$. Secondly, the error in $L^2(\Omega, \rho)$ is determined by the same factors multiplied by the additional constant $D$, which relates the $L^2$-norms over $\rho$ and $\tau$.

**Remark 3.2** The above estimate, which bounds the $L^2$-norm error in terms of a weighted sup-norm, is not quite optimal. It is possible to obtain estimates (in expectation) involving solely $L^2$-norms by slightly modifying the least-squares estimator $\tilde{f}$. For succinctness we shall not do this. See, for instance, [10, 19, 20].

We now move on to the question of optimal sampling. As can be seen in the previous theorem, the samples influence the size of the constant $C$. In the following theorem, we determine a sufficient condition on the sampling measures $\mu_i$ which guarantees that $C \lesssim 1$.

We now make the following standard assumption about the subspace $P \subset L^2(\Omega, \tau)$:

For any $y \in \text{supp}(\tau)$ there exists a $p \in P$ with $p(y) \neq 0$. \hfill (3.8)

Note that this implies that the function $\sum_{i=1}^{N} |\phi_i(y)|^2 > 0$ on $\text{supp}(\tau)$, for any orthonormal basis \{\phi_i\}_{i=1}^{N} of $P$ with respect to $\tau$. In particular, the function

$$ w(y) = \left( \frac{1}{N} \sum_{i=1}^{N} |\phi_i(y)|^2 \right)^{-1}, \quad y \in \text{supp}(\tau), \hfill (3.9) $$

is positive and defined everywhere on $\text{supp}(\tau)$. Notice also that

$$ \int_{\Omega} w^{-1}(y) d\tau(y) = \int_{\Omega} \frac{1}{N} \sum_{i=1}^{N} |\phi_i(y)|^2 d\tau(y) = 1. $$

This function is also independent of the orthonormal basis used. Indeed, $\sum_{i=1}^{N} |\phi_i(y)|^2$ is the reciprocal of the Christoffel function [24] of the subspace $P \subset L^2(\Omega, \tau)$.

**Theorem 3.3.** Let $\gamma, \delta \in (0, 1), P \subset L^2(\Omega, \tau)$ be such that (3.8) holds and \{\phi_i\}_{i=1}^{N} of $P$ with respect to the orthogonality measure $\tau$. Let $w$ be as in (3.9) and suppose that $\mu_1, \ldots, \mu_M$ are probability measures satisfying (3.1) for this choice of $w$. If

$$ M \geq N \log(N/\gamma) ((1 - \delta) \log(1 - \delta) + \delta)^{-1}, \hfill (3.10) $$

then, with probability at least $1 - \gamma$, the constant $C$ defined in (3.6) satisfies $C \leq \frac{1}{\sqrt{1-\delta}}$. Moreover, if $M$ satisfies the slightly stricter condition

$$ M \geq N \log(2N/\gamma) ((1 + \delta) \log(1 + \delta) - \delta)^{-1}, \hfill (3.11) $$

then, with probability at least $1 - \gamma$, the constant $C$ satisfies $C \leq \frac{1}{\sqrt{1-\delta}}$ and the condition number of the matrix $A$ defined by (3.3) satisfies $\kappa(A) \leq \sqrt{\frac{1+\delta}{1-\delta}}$.

Note that the condition (3.11) is stricter than the condition (3.10). We defer the proof of this theorem to §4.
3.3 Choice of sampling measure

Theorem 3.3 implies that $M \gtrsim N \log(N)$ samples are sufficient for a small constant $C$, provided the sampling measures $\mu_i$ are such that

$$\frac{1}{M} \sum_{i=1}^{M} \mu_i(y) = \frac{1}{N} \sum_{i=1}^{N} |\phi_i(y)|^2 \, d\tau(y), \quad \forall y \in \text{supp}(\tau). \quad (3.12)$$

This in turn implies that the least-square estimator $\tilde{f}$ is a quasi-best approximation over $L^2(\Omega, \tau)$ and, provided $D \lesssim 1$, also a quasi-best approximation over $L^2(\Omega, \rho)$. We discuss the constant $D$ in the next section.

Before doing so, let us consider the case $\tau = \rho$, so that $D = 1$. Then one choice of sampling measure that satisfies (3.12) is simply $\mu_1 = \ldots = \mu_M = \mu$, where

$$d\mu(y) = \frac{1}{N} \sum_{i=1}^{N} |\phi_i(y)|^2 \, d\rho(y), \quad y \in \Omega.$$ 

This is the optimal sampling measure introduced in [11]. As noted, a disadvantage of this measure is that it is nonadaptive. If the $N$ increases, the measure $\mu$ changes, and one has to discard the existing samples (in practice, one can recycle at least some of these samples – see [8]). An alternative approach, which avoids this problem, is the following. First, fix $k \in \mathbb{N}$ and let $M = kN$. Then, let

$$d\mu_i(y) = |\phi_j(y)|^2 \, d\rho(y), \quad y \in \Omega, \quad (j-1)k < i \leq jk, \quad j = 1, \ldots, N,$$

so that the first $k$ points are drawn from the measure $|\phi_1|^2 \, d\rho$, the next $k$ points are drawn from the measure $|\phi_2|^2 \, d\rho$ and so forth. Observe that this choice of measures satisfies (3.12):

$$\frac{1}{M} d\mu_i(y) = \frac{1}{M} \sum_{j=1}^{N} k|\phi_j(y)|^2 \, d\rho(y) = \frac{1}{N} \sum_{j=1}^{N} |\phi_j(y)|^2 \, d\rho(y).$$ 

This approach was introduced in [19]. It is clearly adaptive, since if $N$ is incremented by one, we need only sample an additional $k$ points from the new measure $|\phi_{N+1}|^2 \, d\rho$.

3.4 Discrete orthogonality measures

Both of the above approaches require a known orthonormal basis for $P$ and the ability to sample from the corresponding measures in a computationally efficient manner. Neither is typically the case when $\rho$ is a continuous measure on a general domain. To avoid this issue, we now reintroduce the orthogonality measure $\tau$. We construct this as a discrete measure based on a grid $Z = \{z_i\}_{i=1}^{K} \subset \Omega$, where the $z_i$ are independently and identically drawn from the error measure $\rho$. It is worth noting that sampling from $\rho$ may not be trivial in practice. In our experiments, we use rejection sampling. We shall not dwell on this issue any further, since it is domain (and therefore application) specific (see also [6]). We also note that the use of a random grid here is simply to allow one to bound the constant $D$. Deterministic grids are also permitted within this framework, although designing a good grid with $D \lesssim 1$ provably may be nontrivial.

Given such a grid $Z$, we define

$$d\tau(y) = \frac{1}{K} \sum_{i=1}^{K} \delta(y - z_i).$$
We first describe the construction of the orthonormal basis \( \{\phi_1, \ldots, \phi_N\} \) for \( P \subset L^2(\Omega, \tau) \). First, let \( \{\psi_1, \ldots, \psi_N\} \) be a basis for \( P \) in \( L^2(\Omega, \rho) \), and set

\[
B = \left\{ \sqrt{K} \psi_j(z_i) \right\}_{i,j=1}^{K,N} \in \mathbb{C}^{K \times N}.
\]

We assume henceforth that \( B \) is full rank, \( \text{rank}(B) = N \). Note that this is equivalent to the condition \( D < \infty \). Indeed,

\[
Bc = 0 \iff \frac{1}{K} \sum_{i=1}^{K} |p(z_i)|^2 = 0, \quad p = \sum_{i=1}^{N} c_i \psi_i \iff \|p\|_{L^2(\Omega, \tau)} = 0.
\]

In Proposition 3.4 we give a sufficient condition on \( K \) for this to occur. Now let \( B \) have reduced QR decomposition \( B = QR \), where \( Q = \{q_{ij}\} \in \mathbb{C}^{K \times N} \) and \( R \in \mathbb{C}^{N \times N} \). Then it follows straightforwardly that the functions \( \phi_i \) are given by

\[
\phi_i(y) = \sum_{j=1}^{i} (R^{-*})_{ij} \psi_j(y), \quad i = 1, \ldots, N.
\]

### 3.5 Derivation of Methods 1 and 2

We now complete the derivation of Methods 1 and 2. For both methods, we first notice that the function \( w(y) \) defined by (3.9) satisfies

\[
\frac{1}{w(z_i)} = \frac{K}{N} \sum_{j=1}^{N} |q_{ij}|^2, \quad i = 1, \ldots, K. \tag{3.13}
\]

We next consider each method separately:

**Method 1**

We let \( \mu_1 = \ldots = \mu_M = \mu \), where

\[
d\mu(y) = \frac{1}{N} \sum_{i=1}^{N} |\phi_i(y)|^2 \quad d\tau(y) = \frac{1}{K} \sum_{i=1}^{K} \frac{1}{Kw(z_i)} \delta(y - z_i) \quad dy,
\]

and \( w(y) \) is as in (3.9). Let \( \pi = \{\pi_i\}_{i=1}^{K} \) be the probability distribution on \( \{1, \ldots, K\} \) with

\[
\pi_i = \frac{1}{Kw(z_i)} = \frac{1}{N} \sum_{j=1}^{N} |q_{ij}|^2, \quad i = 1, \ldots, K. \tag{3.14}
\]

Then, random sampling \( y \sim \mu \) is effected by randomly choosing an integer \( i \in \{1, \ldots, K\} \) according to \( \pi \) and then setting \( y = z_i \). Let \( i_1, \ldots, i_M \) be \( M \) integers drawn independently from \( \{1, \ldots, K\} \) according to \( \pi \) and \( y_1, \ldots, y_M \) be the sample points. Observe that

\[
A = \left\{ \frac{1}{\sqrt{M}} \sqrt{w(y_j)} \phi_k(y_j) \right\}_{j,k=1}^{M,N} = \left\{ \frac{q_{ik}}{\sqrt{M \pi_i}} \right\}_{j,k=1}^{M,N},
\]

\[
b = \left\{ \frac{1}{\sqrt{M}} \sqrt{w(y_j)} f(y_j) \right\}_{j=1}^{M} = \left\{ \frac{f(z_{ij})}{\sqrt{MK\pi_i}} \right\}_{j=1}^{M}.
\]

This completes the derivation of Method 1.
Method 2

In this case, we fix $M = kN$ for some $k \in \mathbb{N}$ and define, for $(l-1)k < i \leq lk$ and $l = 1, \ldots, N$, the sampling measures

$$d\mu_i(y) = |\phi_l(y)|^2 d\tau(y) = \frac{1}{K} \sum_{i=1}^{K} |\phi_l(z_i)|^2 \delta(y - z_i) dy.$$ 

For each $l$, we define the probability distribution $\pi^{(l)} = \{\pi^{(l)}_i\}_{i=1}^{K}$ on $\{1, \ldots, K\}$ as

$$\pi^{(l)}_i = \frac{1}{K} |\phi_l(z_i)|^2 = |q_{il}|^2, \quad i = 1, \ldots, K.$$ 

Thus, drawing a sample from the $\mu_i$, $(l-1)k < i \leq lk$, is equivalent to $y = z_i$, where $i \sim \pi^{(l)}$. Let $i_1, \ldots, i_M$ be the $M$ integers drawn according to the $\pi^{(l)}$. Then we have

$$A = \left\{ \frac{1}{\sqrt{M}} \sqrt{w(y_j)\phi_k(y_j)} \right\}_{j,k=1}^{M,N} = \left\{ \frac{q_{ij,k}}{\sqrt{\frac{M}{N} \sum_{l=1}^{N} \pi^{(l)}_{ij}}} \right\}_{j,k=1}^{M,N},$$

$$b = \left\{ \frac{1}{\sqrt{M}} \sqrt{w(y_i)f(y_i)} \right\}_{j=1}^{M} = \left\{ \frac{f(z_{ij})}{\sqrt{\frac{MK}{N} \sum_{l=1}^{N} \pi^{(l)}_{ij}}} \right\}_{j=1}^{M}.$$ 

Up to the small modifications needed to make the method adaptive, this completes the derivation of Method 2.

3.6 The size of the grid $Z$

The size $K$ of $Z$ influences the magnitude of the constant $D$. We now estimate this term. For this, we use the following Nikolskii-type inequality for the space $P \subset L^2(\Omega, \rho)$. We let $N(P, \rho)$ be the smallest positive number such that

$$\|p\|_{L^\infty(\Omega)} \leq N(P, \rho) \|p\|_{L^2(\Omega, \rho)}, \quad \forall p \in P.$$  

(3.15)

**Proposition 3.4.** Let $\gamma, \delta \in (0,1)$ and $Z = \{z_i\}_{i=1}^{K}$ where the $z_i$ are drawn independently and identically according to the measure $\rho$ on $\Omega$. If

$$K \geq (N(P, \rho))^2 ((1-\delta) \log(1-\delta) + \delta)^{-1} \log(N/\gamma),$$ 

where $N = \text{dim}(P)$, then with probability at least $1 - \gamma$ the constant $D$ satisfies

$$D \leq \frac{1}{\sqrt{1-\delta}}.$$ 

See [4, Thm. 6.2]. This reduces the question of how large to choose $K$ to that of determining the Nikolskii constant $N(P, \rho)$ for a measure $\rho$ over a domain $\Omega$. As discussed in [4], there are no generic results on this for arbitrary domains and measures. However, in certain cases, one can show that $(N(P, \rho))^2$ is at most quadratic in $N$, the dimension of $P$:

**Definition 3.5** ($\lambda$-rectangle property). A compact domain $\Omega$ has the $\lambda$-rectangle property for some $0 < \lambda < 1$ if it can be written as a (possibly overlapping and uncountable) union $\Omega = \bigcup_{R \in \mathcal{R}} R$ of hyperrectangles $R$ satisfying $\inf_{R \in \mathcal{R}} \text{Vol}(R) = \lambda \text{Vol}(\Omega)$.

See [4, Defn. 6.5]. The following is [4, Thm. 6.6]:
Proposition 3.6. Suppose that $\Omega$ has the $\lambda$-rectangle property and let $P$ be the polynomial space $P = \text{span}\{y \mapsto y^n : n \in \Lambda\}$, where $\Lambda \subset \mathbb{N}^d_0$, $|\Lambda| = N$ is a lower set of multi-indices\footnote{That is, if $n \in \Lambda$ and $n' \leq n$ then $n' \in \Lambda$.}. Let $\rho$ be the uniform probability measure on $\Omega$. Then $(\mathcal{N}(P, \rho))^2 \leq N^2/\lambda$.

We remark in passing that most standard polynomial spaces correspond to lower sets, e.g. tensor product, total degree, hyperbolic cross, and so forth.

Remark 3.7 Unfortunately, while many irregular domains have the the $\lambda$-rectangle property, some simple domains such a balls and simplicies do not \cite{4}. Various of results on the Nikolskii constant (or more generally, the Christoffel function) are known for certain irregular domains, although typically only for total degree polynomial spaces, i.e. those for which $\Lambda = \Lambda_n = \{n = (n_1, \ldots, n_d) : n_1 + \ldots + n_d \leq n\}$. See, for example, \cite{27} for results when $\Omega$ is a ball or simplex, \cite{25} for planar domains with piecewise smooth boundaries, \cite{13} for convex and starlike domains and \cite{12} when $\Omega$ is the surface of the sphere. It is an open problem to determine the Nikolskii constant for more general domains and subspaces $P$.

4 Proofs of the main results

We now prove the main results. The proofs are based on similar ideas to those found in previous works on least-squares approximation. See, for instance, \cite{4,10,11,19}.

Proof of Theorem 3.1 Fix $p \in P$. Then

$$
\|f - \tilde{f}\|^2_{L^2(\Theta, \tau, w)} \leq \|f - p\|^2_{L^2(\Theta, \tau)} + \|\tilde{f} - p\|^2_{L^2(\Theta, \tau)}.
$$

We bound the first term using \cite{3.1} and the fact that the $\mu_i$ are probability measures:

$$
\|f - p\|^2_{L^2(\Theta, \tau)} = \frac{1}{M} \sum_{i=1}^M \int_{\Theta} |f(y) - p(y)|^2 w(y) d\mu_i(y) \leq \sup_{y \in \text{supp}(\tau)} w(y) |f(y) - p(y)|^2.
$$

Hence $\|f - p\|^2_{L^2(\Theta, \tau)} \leq \|f - p\|_{\tau, w}$. For the second term, we first observe that $\langle \tilde{f}, p \rangle_{\tau, w} = \langle f, p \rangle_{\tau, w}$, $\forall p \in P$, since $\tilde{f}$ is a discrete least-squares approximation, and therefore satisfies the normal equations. In particular,

$$
\|\tilde{f} - p\|^2_{\tau, w} = \langle f - p, \tilde{f} - p \rangle_{\tau, w} \leq \|f - p\|_{\tau, w} \|\tilde{f} - p\|_{\tau, w},
$$

and therefore $\|\tilde{f} - p\|_{\tau, w} \leq \|f - p\|_{\tau, w}$. Hence, by the definition of $\mathcal{C}$,

$$
\|\tilde{f} - p\|^2_{L^2(\Theta, \tau)} \leq \mathcal{C} \|\tilde{f} - p\|_{\tau, w} \leq \mathcal{C} \|f - p\|_{\tau, w}.
$$

Furthermore, we have

$$
\|f - p\|^2_{\tau, w} = \frac{1}{M} \sum_{i=1}^M w(y_i) |f(y_i) - p(y_i)|^2 \leq \sup_{y \in \text{supp}(\tau)} w(y) |f(y) - p(y)|^2 = \|f - p\|^2_{\tau, w}.
$$

Combining this with the previous estimate gives $\|\tilde{f} - p\|^2_{L^2(\Theta, \tau)} \leq \mathcal{C} \|f - p\|_{\tau, w}$. Substituting this into \cite{4.1} completes the proof of the first result.
We now consider the second result. We have
\[ \| f - \tilde{f} \|_{L^2(\Omega,\rho)} \leq \| f - p \|_{L^2(\Omega,\rho)} + \| p - \tilde{f} \|_{L^2(\Omega,\rho)} \leq \| f - p \|_{L^2(\Omega,\rho)} + \mathcal{D}\| p - \tilde{f} \|_{L^2(\Omega,\tau)}, \]
\[ \leq \| f - p \|_{L^2(\Omega,\rho)} + \mathcal{C}\mathcal{D}\| p - \tilde{f} \|_{\Upsilon,w}. \]
Thus, using the earlier arguments, we deduce that
\[ \| f - \tilde{f} \|_{L^2(\Omega,\rho)} \leq \| f - p \|_{L^2(\Omega,\rho)} + \mathcal{C}\mathcal{D}\| p - \tilde{f} \|_{\tau,w}, \]
as required.

We now prove Theorem 3.3. For this, we use the following weighted Nikolskii-type inequality for the space \( P \subset L^2(\Omega,\tau) \). For the moment, consider an arbitrary positive function \( w \) defined everywhere on \( \Omega \), and let \( \mathcal{N}(P,\tau,w) \) be the smallest positive number such that
\[ \sup_{z \in \text{supp}(\tau)} \sqrt{w(z)|p(z)|} \leq \mathcal{N}(P,\tau,w)\| p \|_{L^2(\Omega,\tau)}, \quad \forall p \in P. \quad (4.2) \]
Note that the earlier Nikolskii inequality (3.15) is a special case of this weighted inequality, corresponding to \( \tau = \rho \) and \( w \equiv 1 \). At this stage, it is also useful to note the relation between the Nikolskii constant and the Christoffel function of the subspace \( P \subset L^2(\Omega,\tau) \). Specifically, it is straightforward to show that
\[ \mathcal{N}(P,\tau,w) = \sup_{y \in \text{supp}(\tau)} \sqrt{w(y)\sum_{i=1}^{N} |\phi_i(y)|^2}. \quad (4.3) \]
Considering this expression, it becomes clear why \( w(y) \) is taken in (3.9) as proportional to the Christoffel function, since this yields \( \mathcal{N}(P,\tau,w) = \sqrt{N} \). The proof of Theorem 3.3 below relies on this observation.

**Theorem 4.1.** Let \( \gamma,\delta \in (0,1) \) and \( \mu_1,\ldots,\mu_M \) be probability measures satisfying (3.1) for some positive function \( w \) defined almost everywhere on \( \Omega \). If
\[ M \geq (\mathcal{N}(P,\tau,w))^2((1-\delta)\log(1-\delta) + \delta)^{-1}\log(N/\gamma), \]
where \( N = \dim(P) \), then with probability at least \( 1 - \gamma \) the constant \( \mathcal{C} \) defined in (3.6) satisfies \( \mathcal{C} \leq \frac{1}{\sqrt{1-\delta}} \). Moreover, if \( M \) satisfies the slightly stricter condition
\[ M \geq (\mathcal{N}(P,\tau,w))^2((1+\delta)\log(1+\delta) - \delta)^{-1}\log(2N/\gamma), \]
then, with probability at least \( 1 - \gamma \), the constant \( \mathcal{C} \) satisfies \( \mathcal{C} \leq \frac{1}{\sqrt{1+\delta}} \) and the condition number of the matrix \( A \) defined by (3.3) satisfies \( \kappa(A) \leq \sqrt{\frac{1+\delta}{1-\delta}} \).

To prove this result, we first recall the Matrix Chernoff inequality (see [26, Thm. 1.1]):

**Theorem 4.2.** (Matrix Chernoff) Consider a finite sequence \( \{X_k\} \) of independent, random, self-adjoint matrices with dimension \( d \). Assume that each random matrix satisfies
\[ X_k \succeq 0 \quad \text{and} \quad \lambda_{\max}(X_k) \leq R \quad \text{almost surely}. \]
Define
\[ \mu_{\min} := \lambda_{\min}\left(\sum_k \mathbb{E}(X_k)\right) \quad \text{and} \quad \mu_{\max} := \lambda_{\max}\left(\sum_k \mathbb{E}(X_k)\right). \]

Then
\[ \mathbb{P}\left(\lambda_{\min}\left(\sum_k X_k\right) \leq (1 - \delta)\mu_{\min}\right) \leq d \left[ \frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right]^{\mu_{\min}/R}, \quad \forall \delta \in [0, 1], \]
\[ \mathbb{P}\left(\lambda_{\max}\left(\sum_k X_k\right) \geq (1 + \delta)\mu_{\max}\right) \leq d \left[ \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right]^{\mu_{\max}/R}, \quad \forall \delta \geq 0. \]

**Proof of Theorem 4.1** Let \( \{\phi_1, \ldots, \phi_N\} \) be an orthonormal basis of \( P \) with respect to \( \tau, \) \( p \in P, \) \( p \neq 0 \) be arbitrary and write \( p = \sum_{i=1}^N c_i \phi_i, \) so that
\[ \|p\|^2_{L^2(\Omega, \tau)} = \int_{\Omega} \left| \sum_{i=1}^N c_i \phi_i(y) \right|^2 d\tau(y) = \sum_{i=1}^N |c_i|^2 = \|c\|^2_2, \quad c = (c_i)_{i=1}^N. \]
Notice that \( \|p\|^2_{L^2(\Omega, \tau)} = \frac{1}{M} \sum_{i=1}^M w(y_i)|p(y_i)|^2 = c^* Gc, \) where \( G = A^* A \in \mathbb{C}^{N \times N} \) is the self-adjoint matrix with entries \( G_{j,k} = \langle \phi_j, \phi_k \rangle_{\tau, w}. \) It follows that
\[ C = \sup \left\{ \frac{\|c\|}{\sqrt{c^* Gc}} : c \in \mathbb{C}^N, \ c \neq 0 \right\} = \frac{1}{\sqrt{\lambda_{\min}(G)}}, \]
where \( \lambda_{\min}(G) \) is the minimal eigenvalue of \( G. \) Write
\[ G = \sum_{i=1}^M X_i, \quad X_i = \left\{ \frac{1}{M} w(y_i) \phi_j(y_i) \phi_k(y_i) \right\}_{j,k=1}^N. \]
By construction, these matrices are independent and non-negative definite. Also,
\[ (\mathbb{E}(X_i))_{j,k} = \int_{\Omega} \phi_j(y) \overline{\phi_k(y)} w(y) \frac{1}{M} d\mu_i(y), \]
which gives
\[ \left(\sum_{i=1}^M \mathbb{E}(X_i)\right)_{j,k} = \int_{\Omega} \phi_j(y) \overline{\phi_k(y)} w(y) \frac{1}{M} \sum_{i=1}^M d\mu_i(y) = \int_{\Omega} \phi_j(y) \overline{\phi_k(y)} d\tau(y) = \delta_{j,k}. \]
Hence \( \sum_{i=1}^M \mathbb{E}(X_i) = I \) is the identity matrix. Moreover, for any \( c \in \mathbb{C}^N \) we have
\[ c^* X_i c = \frac{1}{M} \sum_{j=1}^N c_j \sqrt{w(y_i)} \phi_j(y_i) \leq \frac{(\mathbb{N}(P, \tau, w))^2 M}{\sum_{i=1}^N c_j^2} \|c\|^2_{L^2(\Omega, \tau)} = \frac{(\mathbb{N}(P, \tau, w))^2}{M} \|c\|^2_2. \]
Since these matrices are self-adjoint and non-negative definite, we deduce that
\[ \lambda_{\max}(X_i) \leq \frac{(\mathbb{N}(P, \tau, w))^2}{M}. \]
We now apply the Matrix Chernoff bound with \( d = N, R = (N(P, \tau, w))^2 / M \) and

\[
\mu_{\min} = \lambda_{\min} \left( \sum_{m=1}^{M} \mathbb{E}(X_m) \right) = \lambda_{\min} (I) = 1,
\]
to get

\[
\mathbb{P} \left( C \geq \frac{1}{\sqrt{1 - \delta}} \right) = \mathbb{P} \left( \lambda_{\min} \left( \sum_{m=1}^{M} X_m \right) \leq (1 - \delta) \right) \leq N \left[ \frac{e^{-\frac{\delta}{(1 - \delta)^{1 - \delta}}}}{(1 - \delta)^{1 - \delta}} \right]^{1/R},
\]

\[
= N \exp \left( -\frac{(1 - \delta) \log(1 - \delta) + \delta}{M^{-1}(N(P, \tau, w))^2} \right).
\]
The condition on \( M \) implies that \( \mathbb{P} \left( C \geq \frac{1}{\sqrt{1 - \delta}} \right) \leq \gamma \), which gives the first result.

For the second result, we note that \( \kappa(A) = \sqrt{\lambda_{\max}(G) / \lambda_{\min}(G)} \). Hence, by the Matrix Chernoff bound with \( d, N \) and \( \mu_{\min} \) as above and \( \mu_{\max} = 1 \), we have

\[
\mathbb{P} \left( \kappa(A) \geq \sqrt{\frac{1 + \delta}{1 - \delta}} \right) \leq \mathbb{P} \left( \lambda_{\min}(G) \leq (1 - \delta) \right) + \mathbb{P} \left( \lambda_{\max}(G) \geq (1 + \delta) \right)
\]

\[
\leq N \left( \exp \left( -\frac{(1 - \delta) \log(1 - \delta) + \delta}{M^{-1}(N(P, \tau, w))^2} \right) + \exp \left( -\frac{(1 + \delta) \log(1 + \delta) - \delta}{M^{-1}(N(P, \tau, w))^2} \right) \right).
\]

Note that \( (1 + \delta) \log(1 + \delta) - \delta \leq (1 - \delta) \log(1 - \delta) + \delta \) for \( 0 < \delta < 1 \). Hence

\[
\mathbb{P} \left( \kappa(A) \geq \sqrt{\frac{1 + \delta}{1 - \delta}} \right) \leq 2N \exp \left( -\frac{(1 + \delta) \log(1 + \delta) - \delta}{M^{-1}(N(P, \tau, w))^2} \right) \leq \gamma,
\]
where in the last step we use the condition on \( M \). This completes the proof. \( \square \)

**Proof of Theorem 3.3** The result follows from (4.3) the definition of \( w \) (3.9). Indeed, we have \( N(P, \tau, w) = \sqrt{N} \) for this choice of \( w \). Hence Theorem 4.1 gives the result. \( \square \)

We conclude this section with the proofs of Theorems 2.1 and 2.2.

**Proof of Theorem 2.1** Theorem 3.3 and the condition on \( M \) imply that \( C \leq 1/\sqrt{1 - \delta} \) and \( \kappa(A) \leq \sqrt{1 + \delta} / \sqrt{1 - \delta} \) with probability at least \( 1 - \gamma / 2 \), and Proposition 3.4 and the condition on \( K \) imply that \( D \leq 1/\sqrt{1 - \delta} \) with probability at least \( 1 - \gamma / 2 \). Hence \( C \leq 1/\sqrt{1 - \delta}, D \leq 1/\sqrt{1 - \delta} \) and \( \kappa(A) \leq \sqrt{1 + \delta} / \sqrt{1 - \delta} \) with probability at least \( 1 - \gamma \). The condition on \( D \) implies that \( B \) is full rank (see (3.4)). Next, observe that (3.13) and (3.14) give

\[
\|g\|_{\tau,w} = \sup_{y \in \text{supp}(\tau)} \sqrt{w(y)} |g(y)| = \max_{i=1,\ldots,K} \sqrt{w(z_i)} |g(z_i)| = \max_{i=1,\ldots,K} \left\{ \frac{|g(z_i)|}{\sqrt{K\pi_i}} \right\} = \|g\|_{Z,\pi}.
\]
The result now follows from Theorem 3.1. \( \square \)

**Proof of Theorem 2.2** For each \( t \), let \( C = C_t \) and \( D = D_t \) be the corresponding constants and write \( A = A_t, B = B_t \) for the matrices defined in Step 7 and Step 4 of Method 2 respectively. Define
the following events:

\[ E : D_r \leq \frac{1}{\sqrt{1 - \delta}}, \]

\[ F_t : C_t \leq \frac{1}{\sqrt{1 - \delta}} \quad \text{and} \quad \kappa(A_t) \leq \sqrt{\frac{1 + \delta}{1 - \delta}}, \quad t = 1, \ldots, r, \]

\[ G = E \cap F_1 \cap \ldots \cap F_r. \]

Suppose first that event \( G \) occurs. Notice that \( D_1 \leq D_2 \leq \ldots \leq D_r \leq 1/\sqrt{1 - \delta} \) since the \( P_t \) are nested subspaces. Hence \( B_t \) is full rank for every \( t \), which gives (i). Also, as in the proof of the previous theorem, the events \( F_t \) imply (ii) and the events \( E \) and \( F_t \) imply (iii).

It remains to show \( \mathbb{P}(G) \geq 1 - \gamma \). By the union bound \( \mathbb{P}(G^c) \leq \mathbb{P}(E^c) + \mathbb{P}(F_1^c) + \ldots + \mathbb{P}(F_r^c) \). Proposition 3.4 and the condition on \( K \) give that \( \mathbb{P}(E^c) \leq \gamma/2 \). Moreover, since \( M_t = k_t N_t \), Theorem 3.3 and the condition on \( k_t \) give that \( \mathbb{P}(F_t^c) \leq \gamma_t/2 \). Hence \( \mathbb{P}(G^c) \leq \gamma/2 + \sum_{t=1}^{r} \gamma_t/2 = \gamma \), as required.

5 Numerical examples

To conclude this paper, we demonstrate Methods 1 and 2 on several examples. Throughout, we consider the approximation of smooth functions using polynomials. In particular, we choose \( P = P_n^{HC} = \text{span} \{ y \mapsto y^n : n \in \Lambda_n^{HC} \} \), where \( \Lambda_n^{HC} \) is the hyperbolic cross index set of index \( n \):

\[ \Lambda_n^{HC} = \left\{ n = (n_1, \ldots, n_d) \in \mathbb{N}_0^d : \prod_{k=1}^{d} (n_k + 1) \leq n + 1 \right\}. \]

We take \( \Omega \) to be a compact domain contained in the unit hypercube \([-1,1]^d\), and \( \rho \) to be the uniform measure on \( \Omega \). Sampling from \( \rho \) is performed by rejection sampling. The initial basis \( \{ \psi_1, \ldots, \psi_N \} \) for \( P \) is constructed by taking the restrictions to \( \Omega \) of the orthonormal Legendre polynomials on \([-1,1]^d\) with indices belonging to \( \Lambda \). This approach is based on [4]. Note, however, that [4] does not seek to orthogonalize this basis, unlike the approach developed in this paper.

In our experiments we first generate a grid of size \( K = 20000 \) points, and then compute the approximation for values \( 1 \leq N_1 < \ldots < N_r \leq 1000 \). In the first series of experiments, we compute the relative approximation error

\[ E_t(f) = \frac{\| f - \tilde{f} \|_{L^2(\Omega, \tau)}}{\| f \|_{L^2(\Omega, \tau)}}, \quad (5.1) \]

over this grid, as well as the constant \( C \). Later, in Fig. 4 we consider an error computed over a distinct grid of points. The computed values of \( E_t(f) \) and \( C \) values are averaged over 50 trials, as follows. For Method 1, for each \( N \) we take 50 independent draws of the corresponding \( M \) sample points \( \{ y_i \}_{i=1}^{M} \) and take the mean values of \( C \) and the error. For Method 2, we perform 50 independent experiments sweeping, as described in the method, over \( N_1, \ldots, N_r \) and then take the mean values.

To examine the benefit of the new sampling procedure, we also consider uniform random sampling over the fine grid of \( K \) points (this type of sampling is similar to that used [4], although not identical, since [4] considers random sampling with respect to the continuous uniform measure.
on Ω). This method is referred to as ‘Uniform’ in our experiments. For functions, we use the following:

\[ f_1(y) = \exp \left( -\sum_{i=1}^{d} y_i/d \right), \quad f_2(y) = \frac{1}{\sum_{i=1}^{d} \sqrt{|y_i|}}, \]

\[ f_3(y) = \prod_{i=1}^{d} \frac{d/4}{d/4 + (y_i + (-1)^{i+1}/(i + 1))^2}, \quad f_4(y) = \frac{1}{y_1^2 + y_2^2}. \]

Note that \( f_3 \) is known as the ‘Genz product peak’ function.

In Fig. 1 we compare Method 1 and Method 2 with uniform random sampling (Uniform) over three domains in various different dimensions. The domains (for \( d = 2 \)) are shown in Fig. 2 along with the fine grid of \( K \) points and the samples generated by Method 1 for a typical value of \( M \). In all cases, both procedures lead to an improvement over uniform sampling. The effect is most noticeable for lower dimensions; such an observation is not surprising, given that in higher dimensions the maximum degree \( n \) of the hyperbolic cross index set \( \Lambda^\text{HC}_n \) is not as large as in lower dimensions (in our experiments, \( N = |\Lambda^\text{HC}_n| \leq 1000 \)). For Uniform the error actually increases with \( N \) in some cases. This effect is most dramatic for \( f_2 \) which, unlike the other functions, is not infinitely smooth in Ω. The reason for this increase is because the number of samples \( M \approx N \log(N) \) in this experiment, which is asymptotically lower order than the quadratic scaling \( N^2 \log(N) \) known to be sufficient (for certain domains, see [4]) when working with uniform random samples.

This experiment also demonstrates that Method 1 and Method 2 have similar performance in all cases. Recall, however, that Method 1 recycles none of its samples when \( N \) increases, whereas Method 2 recycles all its samples.

In Fig. 3 we examine the constant \( C \) for the three domains and methods, and for different scalings of \( M \) with \( N \). For Uniform, all scalings lead to an exponentially increasing constant \( C \) – a well-known phenomenon [7]. For Method 1 and Method 2, notice that any linear scaling \( M = cN \) eventually leads to a growing constant \( C \), whereas \( C \) remains bounded for either of the two log-linear scalings \( M = cN \log(N) \). This result therefore verifies Theorem 3.3.

Remark 5.1 As discussed in [4] (see also [5, 6]), the matrix \( B \) is in general ill-conditioned since the initial basis \( \{\psi_1, \ldots, \psi_N\} \) is nearly redundant for large \( N \). Therein, this is tackled by using a truncated SVD. Conversely, this ill-conditioning is seemingly not an issue in Methods 1 and 2. As shown in [20], an explanation for this is that the ill-conditioning of \( B \) does not prohibit generating a near-orthonormal basis for \( P \) over the \( K \)-grid. Of course, if such a grid cannot be generated (see [6]) then one may have no choice but to work with the basis \( \{\psi_1, \ldots, \psi_N\} \) and the resulting ill-conditioning, as in [4].

Thus far, the error has been computed via (5.1) over the same grid of \( K \) points used in the method (this corresponds to the first error bound in Theorem 3.1). It is also important to consider the approximation’s performance off this grid. In Fig. 4 we consider the error

\[ E_{\tilde{\tau}}(f) = \frac{\|f - \tilde{f}\|_{L^2(\Omega, \tilde{\tau})}}{\|f\|_{L^2(\Omega, \tilde{\tau})}}, \quad (5.2) \]

where \( \tilde{\tau} \) is a discrete measure over \( T \) points, chosen randomly from the uniform measure on \( \Omega \) and independently of the points used in the \( K \)-grid \( \tau \). As this figure shows, when \( K = 20000 \), the error \( E_{\tilde{\tau}}(f) \) behaves significantly worse than the error \( E_{\tau}(f) \), which is shown in top row of Fig. 1. In view of Theorem 3.1 this indicates that the \( K \)-grid is not large enough to ensure the constant \( D \) is
Figure 1: The error $E(f)$ versus $M$, with $M$ chosen as the smallest value such that $M \geq N \log(N)$. First row: the domain $\Omega_1 = \{y : 1/4 \leq \|y\|_2 \leq 1\}$ with $f = f_1$. Second row: $\Omega_1$ and $f = f_2$. Third row: the domain $\Omega_2 = \{y \in (-1,1)^d : y_1 + \ldots + y_d \leq 1\}$ with $f = f_3$. Fourth row: the domain $\Omega_3 = \{y \in (-1,1)^d : y_1^2 + y_2^2 \geq 1/4\}$ with $f = f_4$. The domains are shown in Fig. 2 for $d = 2$. 
Figure 2: The domains $\Omega_1$, $\Omega_2$ and $\Omega_3$ (left to right) used in Fig. 1 for $d = 2$, the fine grid with $K = 20000$ points, and the samples generated in a typical instance of Method 1 with $M = 1000$ and $N = 198$.

Figure 3: The constant $C$ versus $N$ in $d = 2$ dimensions for the domains $\Omega_1$, $\Omega_2$, $\Omega_3$ (top to bottom) used in Fig. 1.
Figure 4: The error $E_{\tau}(f)$ versus $M$, with $M$ chosen as the smallest value such that $M \geq N \log(N)$, for
$\Omega_1 = \{ \mathbf{y} : 1/4 \leq \| \mathbf{y} \|_2 \leq 1 \}$ and $f = f_1$. Top row: $T = K = 20000$. Middle row: $T = K = 40000$. Bottom row: $T = K = 80000$. 

small. As expected, doubling $K$ yields better behaviour for the error $E_\tau(f)$, and doubling it once more yields a further improvement.

This experiment demonstrates one of the challenges with approximation over general domains. Namely, while it is straightforward to check for good accuracy over the $K$-grid (indeed, one simply computes $C$), it is difficult to ensure a priori good performance over the whole of $\Omega$. Indeed, the constant $D$ is not straightforwardly computable. As noted, if $\Omega$ has the $\lambda$-rectangle property, then one may use Propositions 3.4 and 3.6 to estimate $K$ (we caution, however, this estimate may not be particularly useful if $\lambda$ is small). But for domains not satisfying this property, as is the case for the domain considered in Fig. 4, the answer to the question of how large $K$ should be to ensure good accuracy over $\Omega$ is currently unknown.

Finally, in Fig. 5 we demonstrate a curious phenomenon: in certain cases, there may be far less benefit from using these methods over uniform sampling. In this experiment we consider an annular domain with outer radius $1/2$. In this case, quite in contrast to what was seen for the annular domain $\Omega_1$ (which has outer radius 1) in Fig. 1, the approximation converges. Furthermore, neither Method 1 nor Method 2 achieves a better rate of convergence. Fig. 6 shows the constant $C$ for all three methods when $d = 2$. Unlike in the case of $\Omega_1$ (see Fig. 3) for Uniform the constant $C$ remains bounded when log-linear sampling is used, although it is several orders of magnitude larger than for Methods 1 and 2 with the same scaling. This phenomenon relates to the fact that the domain $\Omega$ in this case is compactly contained in $(-1,1)^d$. Hence the Legendre polynomials on $[-1,1]^d$, when restricted to $\Omega$ constitute a frame. See [4, Sec. 8.2] for further discussion.
6 Conclusions

In this paper, we introduced new methods for optimal weighted least-squares approximations in general domains and arbitrary finite-dimensional spaces based on a discrete grid $Z$. As we showed, under the log-linear sample complexity $M \gtrsim N \log(N)$, these methods are provably well conditioned and accurate over the grid. This improves on the method introduced in [4], for which the sampling complexity at best quadratic in $N$. To obtain accuracy over the original space, a random grid is used, whose size $K$ is related to the Nikolskii constant $\mathcal{N}(P, \rho)$. For domains possessing the so-called $\lambda$-rectangle property the quadratic scaling $K \gtrsim N^2 \log(N)$ is sufficient. We introduce two versions of this method, with the latter recycling all its samples when the approximation $P$ is augmented. Numerical experiments confirm the benefits of these methods over standard sampling.

As noted in the previous section, a shortcoming of this method is ensuring $K$ is large enough to guarantee a good approximation over $\Omega$ when $\Omega$ does not satisfy the $\lambda$-rectangle property. Obtaining estimates for general domains is an open problem. Another limitation is that it must be possible to sample from the continuous measure $\rho$ over $\Omega$ in order to generate the $K$-grid. In particular, and unlike in the method introduced in [4], $\Omega$ must be known in advance in some suitable sense, and even then, procedures such as rejection sampling may become prohibitively expensive in high dimensions. In practice, $\Omega$ may not be known in advance: see [4], for instance, for examples motivated by uncertainty quantification where $\Omega$ can only be ‘learned’ as the samples of $f(y)$ are taken. An objective of future work is to investigate whether the procedure developed in this paper can be extended to certain settings where generating a fine grid beforehand is not feasible.

Acknowledgments

This work was supported by the PIMS CRG in “High-dimensional Data Analysis” and by NSERC through grant R611675.

References

[1] B. Adcock. Infinite-dimensional compressed sensing and function interpolation. Found. Comput. Math. 18(3):661–701, 2018.

[2] B. Adcock, A. Bao, and S. Brugiapaglia. Correcting for unknown errors in sparse high-dimensional function approximation. Numer. Math. 142(3):667–711, 2019.

[3] B. Adcock, S. Brugiapaglia, and C. G. Webster. Compressed sensing approaches for polynomial approximation of high-dimensional functions. In Compressed Sensing and Its Applications. Birkhäuser, 2017.

[4] B. Adcock and D. Huybrechs. Approximating smooth, multivariate functions on irregular domains. arXiv:1802.00602, 2018.

[5] B. Adcock and D. Huybrechs. Frames and numerical approximation II: generalized sampling. arXiv:1802.01950, 2018.

[6] B. Adcock and D. Huybrechs. Frames and numerical approximation. SIAM Rev., 61(3):443–473, 2019.

[7] B. Adcock, R. Platte, and A. Shadrin. Optimal sampling rates for approximating analytic functions from pointwise samples.IMA J. Num. Anal. (to appear), 2018.

[8] B. Arras, M. Bachmayr, and A. Cohen. Sequential sampling for optimal weighted least squares approximations in hierarchical spaces. SIAM J. Math. Data Sci., 1(1):189–207, 2019.
[9] A. Chkifa, N. Dexter, H. Tran, and C. G. Webster. Polynomial approximation via compressed sensing of high-dimensional functions on lower sets. *Math. Comp.*, 2017.

[10] A. Cohen, M. A. Davenport, and D. Leviatan. On the stability and accuracy of least squares approximations. *Found. Comput. Math.*, 13:819–834, 2013.

[11] A. Cohen and G. Migliorati. Optimal weighted least-squares methods. *SMAI Journal of Computational Mathematics*, 3:181–203, 2017.

[12] F. Dai, D. Gorbachev, and S. Tikhonov. Estimates of the asymptotic Nikolskii constants for spherical polynomials. *arXiv:1907.03832*, 2019.

[13] A. Doostan and H. Owhadi. A non-adapted sparse approximation of PDEs with stochastic inputs. *J. Comput. Phys.*, 230(8):3015–3034, 2011.

[14] J. Hampton and A. Doostan. Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies. *J. Comput. Phys.*, 280:363–386, 2015.

[15] A. Kroó. Christoffel functions on convex and starlike domains in $\mathbb{R}^d$. *J. Math. Anal. Appl.*, 421:718–729, 2015.

[16] C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems*, volume 15. SIAM, 1995.

[17] G. Migliorati. Polynomial approximation by means of the random discrete $L^2$ projection and application to inverse problems for PDEs with stochastic data. PhD thesis, Politecnico di Milano, 2013.

[18] G. Migliorati. Multivariate Markov-type and Nikolskii-type inequalities for polynomials associated with downward closed multi-index sets. *J. Approx. Theory*, 189:137–159, 2015.

[19] G. Migliorati. Adaptive approximation by optimal weighted least squares methods. *arXiv:1807.00402*, 2018.

[20] G. Migliorati. Multivariate approximation of functions on irregular domains by weighted least-squares methods. *arXiv:1907.12304*, 2019.

[21] G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone. Analysis of the discrete $L^2$ projection on polynomial spaces with random evaluations. *Found. Comput. Math.*, 14:419–456, 2014.

[22] A. Narayan, J. D. Jakeman, and T. Zhou. A Christoffel function weighted least squares algorithm for collocation approximations. *arXiv:1412.4305*, 2014.

[23] A. Narayan and T. Zhou. Stochastic collocation on unstructured multivariate meshes. *Commun. Comput. Phys.*, 18(1):1–36, 2015.

[24] P. Nevai. Géza Freud, orthogonal polynomials and Christoffel functions. A case study. *J. Approx. Theory*, 48(1):3–167, 1986.

[25] A. Prymak and O. Usoltseva. Christoffel function on planar domains with piecewise smooth boundary. *Acta Math. Hungar.*, 158(1):216–234, 2019.

[26] J. A. Tropp. User friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12:389–434, 2012.

[27] Y. Xu. Orthogonal polynomials and cubature formulae on spheres and on balls. *SIAM J. Math. Anal.*, 29(3):778–793, 1998.