Approximating smooth, multivariate functions on irregular domains

Ben Adcock∗ Daan Huybrechs†

February 5, 2018

Abstract

In this paper, we analyze a method known as polynomial frame approximation for approximating smooth, multivariate functions defined on irregular domains in $d$ dimensions, where $d$ can be arbitrary. This method is simple, and relies only on orthogonal polynomials on a bounding tensor-product domain. In particular, the domain of the function need not be known in advance. When restricted to a subdomain, an orthonormal basis is no longer a basis, but a frame. Numerical computations with frames present potential difficulties, due to the near-linear dependence of the finite approximation system. Nevertheless, well-conditioned approximations can be obtained via regularization; for instance, truncated singular value decompositions. We comprehensively analyze such approximations in this paper, providing error estimates for functions with both classical and mixed Sobolev regularity, with the latter being particularly suitable for higher-dimensional problems. We also analyze the sample complexity of the approximation for sample points chosen randomly according to a probability measure, providing estimates in terms of the Nikolskii-type inequality for the domain. For a large class of nontrivial domains, we show that the sample complexity for points drawn from the uniform measure is quadratic in the dimension of the polynomial space, independently of $d$.

1 Introduction

Many problems in scientific computing call for the approximation of smooth, multivariate functions. This problem is often challenging, due to the curse of dimensionality. Yet significant strides have been made over the last several decades towards mitigating this problem, typically by assuming some anisotropic behaviour of the function to approximate. Approaches such as sparse grids [15] have enjoyed substantial success in the numerical solution of high-dimensional PDEs, and more recently techniques based on computing multivariate polynomial approximations – often referred to as generalized polynomial chaos expansions [45] – have begun to be commonly used for problems in Uncertainty Quantification (UQ) (see [4, 17, 18, 22, 21, 26, 35, 47] and references therein).

The majority of algorithms for high-dimensional approximation assume the underlying function $f$ is defined over a tensor-product domain. The key benefit of doing so is simplicity. For instance, the orthogonal polynomials on a tensor-product domain with respect to a tensor-product measure are precisely tensor products of the corresponding one-dimensional orthogonal polynomials. Yet there are many practical instances where the domain of interest is not of tensor-product type. A particular instance is surrogate model construction in UQ. In practice, it is often the case that the random variables are correlated [44], which leads to an irregular domain. Alternatively or in addition, the
given forward model may not be well-defined over the whole of the assumed tensor-product domain, or may produce unphysical values in certain regions (e.g. negative volumes). This in effect leads to failed evaluations, resulting once more in an irregular domain. Similarly, in the field of model order reduction, techniques such as active subspaces naturally lead to approximation problems over irregular domains. For example, when a function defined on a high-dimensional hypercube is projected to a function of a reduced set of parameters, the resulting domain (the projection of the hypercube) is generally polyhedral; a so-called zonotope. Finally, many applications in UQ also involve forward models which are piecewise smooth (see and references therein). Unless such discontinuities happen to be aligned along coordinate axes, this effectively results in approximation problems involving two or more smooth functions defined over irregular domains.

With this issue in mind, the purpose of this paper is to present a systematic study of a simple but effective technique for approximating high-dimensional functions defined on irregular domains. It is based on using tensor-product orthogonal polynomials on a bounding box, and is referred to as polynomial frame approximation. The approach corresponds to approximation in a frame, rather than a basis, since there are potentially many ways the unknown function on the irregular domain can be represented in a basis on the bounding box. Our main results demonstrate that this procedure achieves (to a significant degree) the four primary criteria for a numerical approximation scheme: namely, simplicity, accuracy, stability and efficiency. We define these terms formally in the next section, however we note in passing that simplicity means that the same procedure can be applied to virtually any irregular domain, with only minor modifications. In particular, no costly parametrization of the domain or its boundary (a potentially infeasible task in high dimensions) is required in the construction of the approximation.

A main contribution of this paper is the rigorous analysis of polynomial frame approximations. Central to this is the notion of frames of Hilbert spaces, as opposed to more conventional orthogonal bases. We stress at this point that our technique does not attempt to orthogonalize a basis. Instead, it relies on the particular properties of frames to achieve accurate and stable approximations. A key element of frame approximations (not just of polynomial type) is that they lead to highly ill-conditioned linear systems of equations. However, using regularization we are able to obtain a mapping from the sample points to the polynomial space that is both well-conditioned and accurate. Crucially, we determine approximation rates and sample complexity estimates that scale well with the underlying dimension, thus mitigating the curse of dimensionality.

Before proceeding further, it is worth noting that polynomial frame approximation, and variations thereof, are in essence already used in the aforementioned applications. Indeed, any approach to UQ which computes a generalized polynomial chaos expansion from function evaluations which are limited (due to the particular problem at hand) to a non-tensorial subdomain is completely equivalent to polynomial frame approximation. See for further details. However, a thorough analysis of the accuracy, stability and efficiency of such approximations – in particular, exploiting the connections to frame theory as we do in this paper – is, to the best of our knowledge, lacking.

Besides providing the first clear theoretical explanation for why these algorithms work in practice, we also expect the results of this paper to also shed light on ways in which to improve them. For example, the problem of designing better sampling sets for irregular domains; a topic of significant practical interest.

2 Overview of the paper

In this section, we give a short overview of the main aspects of the paper.
2.1 Polynomial frame approximations

This paper concerns the approximation of a smooth multivariate function \( f : \Omega \to \mathbb{C} \) defined over a non-tensor product domain \( \Omega \subset \mathbb{R}^d \). The approximation is based on four key steps:

(i) Choose a tensor-product domain \( D \) such that \( \Omega \subseteq D \).

(ii) Choose a tensor-product probability measure \( \nu \) on \( D \), a tensor-product orthonormal basis \( \{ \psi_n \}_{n} \) of \( L^2(D, \nu) \) and a finite index set \( \Lambda \) with \( |\Lambda| = N \).

(iii) Take \( M \) samples of \( f \) of the form \( f(y_1), \ldots, f(y_M) \) where \( \Upsilon = \{ y_1, \ldots, y_M \} \subset \Omega \).

(iv) Compute an approximation to \( f \) of the form \( f_{\Upsilon, \Lambda} = \sum_{n \in \Lambda} c_n \phi_n \), where \( \phi_n = \psi_n|_{\Omega} \).

This immediately raises a number of questions, discussed next:

1. How to compute the approximation. There are several options for doing this, including interpolation if \( M = |\Lambda| = N \), sparse regularization (i.e. compressed sensing) if \( M < N \) and least-squares fitting if \( M > N \). We shall primarily consider the latter. Interpolation requires good choices of interpolation nodes \( y_1, \ldots, y_N \) so as to maintain small Lebesgue constants, and it is unclear how to design such nodes for general irregular domains. Compressed sensing is an interesting option, however beyond the scope of this paper (see \( \S 9 \) for some further discussion). Least-squares fitting, on the other hand, is a popular tool for high-dimensional approximation on tensor-product domains \([17, 24, 31, 33, 34, 35, 48]\), and has the twin benefits of being simple to implement and analyze.

Note that the least-squares approximation \( f_{\Upsilon, \Lambda} \) is defined precisely as

\[
    f_{\Upsilon, \Lambda} = \arg\min_{p \in P_\Lambda} \frac{1}{M} \sum_{y \in \Upsilon} |f(y) - p(y)|^2, \tag{2.1}
\]

where \( P_\Lambda = \text{span}\{\phi_n : n \in \Lambda\} \) is the finite-dimensional approximation space. Equivalently, the coefficients \( c = (c_n)_{n \in \Lambda} \) of \( f_{\Upsilon, \Lambda} \) are the solution of the algebraic least-squares problem

\[
    c = \arg\min_{x \in \mathbb{C}^N} \| Ax - b \|_2, \tag{2.2}
\]

where \( A = \left\{ \frac{1}{\sqrt{M}} \phi_n(y) \right\}_{y \in \Upsilon, n \in \Lambda} \in \mathbb{C}^{M \times N} \) and \( b = \left\{ \frac{1}{\sqrt{M}} f(y) \right\}_{y \in \Upsilon} \in \mathbb{C}^M \).

2. How to choose the orthonormal basis \( \{ \psi_n \}_{n} \) and index set \( \Lambda \). These choices determine the accuracy of the approximation. Smooth functions are typically well-approximated by polynomials, so we shall generally take \( \{ \psi_n \}_{n} \) to be an orthonormal tensor-product polynomial basis. Our main numerical examples will consider tensor-product Legendre polynomials. We will also highlight the possibility of nonpolynomial approximations, for example using a cosine basis when \( \Omega \) is compactly contained in \( D = (-1,1)^d \). Given the basis \( \{ \psi_n \}_{n} \), we shall consider several standard choices for \( \Lambda \), including total degree and hyperbolic cross index sets, or more generally, so-called lower sets. These sets have been studied quite extensively for polynomial approximations in tensor-product domains (see \([3, 17, 18, 19, 23, 32, 33]\) and references therein).

3. How to choose the sample points \( \Upsilon \). Our primary concern in this regard lies with the sampling efficiency (or sample complexity) of the approximation: namely, how large \( M \) must be in relation to \( N = |\Lambda| \) to ensure a quasi-optimal approximation. The problem of designing optimal sampling points for high-dimensional polynomial approximation remains open even in tensor-product
domains (although we note in passing some recent constructions which require only \( M = O(N) \) sample points [24]). We shall therefore not attempt to solve it for irregular domains. Instead, we consider straightforward random samplings. Specifically, we draw \( y_1, \ldots, y_M \) independently according to a suitable probability measure on \( \Omega \) (for example, the uniform measure whenever \( \Omega \) is compact). This approach, although simple, permits concrete sample complexity estimates for a large class of domains \( \Omega \) which are quadratic in \( N = |\Lambda| \) for any dimension \( d \). Up to domain-dependent constants (which we determine), this quadratic sample complexity is the same as the corresponding result for compact tensor-product domains when the sample points are drawn from the uniform measure [17]. This scaling is known to be essentially sharp.

2.2 Conditioning and stability

Polynomial frame approximation is certainly simple, and it is tempting to think that it can achieve high accuracy with relatively good efficiency. After all, the method computes a polynomial approximation in a domain, albeit an irregular one. Unfortunately, the matrix \( A \) of the algebraic least-squares problem (2.2) is extremely ill-conditioned, even when \( M \gg N \) (we estimate this ill-conditioning later in the paper for relevant examples). This is due to the fact that the set \( \{ \phi_n \}_n \) is not a basis for \( L^2(\Omega, \mu) \), but rather a frame (see §3.3 for the definition of a frame). Frames are typically redundant, meaning that any function \( f \) has infinitely-many expansions of the form \( f = \sum_n c_n \phi_n \) with coefficients \( \{ c_n \}_n \) in \( \ell^2 \). When translated to the finite setting, this redundancy means that the truncated Gram matrix

\[
G_{\Lambda} = \{ \langle \phi_m, \phi_n \rangle_{L^2(\Omega, \mu)} \}_{m,n \in \Lambda},
\]

which is a finite approximation to the singular infinite Gram matrix, is typically extremely poorly conditioned for large \( N \) [5]. Note that \( \mathbb{E}(A^*A) = G_{\Lambda} \) if the sample points \( y_i \) are drawn according to a suitable measure on \( \mu \) (see (3.2)); hence the least-squares matrix \( A \) is expected to inherit similar ill-conditioning.

In the face of such an ill-conditioned least-squares problem, one would usually expect it to be impossible to achieve high accuracy with a standard numerical implementation. However, this expectation turns out to be incorrect. The frame property endows the problem with sufficient structure so that accurate, well-conditioned approximations can be computed via a simple regularization procedure. We show in this paper that regularized least-squares solutions (computed via hard thresholding of the singular values of \( A \)) yield well-conditioned approximations which converge rapidly down to the thresholding parameter \( \epsilon \) (typically set according to some desired level of accuracy). Besides leading to provably desirable properties, the inclusion of a threshold parameter is also often convenient in practice. Since function samples in the aforementioned applications are always contaminated with numerical error, this parameter can be tuned in relation to the error to obtain a better approximation.

Remark 2.1 We stress that the frame property is crucial in endowing the approximation with stability and accuracy, hence why we refer to this approach as \textit{polynomial frame approximations}. Choosing \( \{ \phi_n \}_n \) to be the monomial basis also leads to an exceedingly ill-conditioned problem, but one with no expectation of numerical stability. The underlying reason for this is that the frame property guarantees existence of expansions \( f = \sum_n c_n \phi_n \) in the frame for which the coefficients \( \{ c_n \}_n \) decay rapidly (accuracy) and have bounded \( \ell^2 \)-norm (stability). This is not the case for the monomial basis, except under very restrictive conditions on \( f \). See [5] [6] for further discussion.

\footnote{As discussed in [8] (see also [38]), in one dimension if the sample points are deterministic and exactly equispaced, then the least-squares approximation is unstable unless the number of sample points scales quadratically in the polynomial degree \( N \).}
2.3 Main results

We now summarize our main results.

**Accuracy and conditioning.** Our first result concerns the accuracy and condition number of the regularized least-squares approximation. As mentioned above, we compute this via a truncated SVD of the least-squares matrix $A$ using a threshold parameter $\epsilon > 0$. Write $f_{T,\Lambda,\epsilon}$ for this approximation and $c^\epsilon$ for its coefficients in the system $\{\phi_n\}_{n \in \Lambda}$.

**Theorem 2.2** (Accuracy and conditioning). There exists a constant $C_{T,\Lambda,\epsilon} > 0$ such that

$$\|f - f_{T,\Lambda,\epsilon}\|_{L^2(\Omega,\mu)} \leq (1 + C_{T,\Lambda,\epsilon}) E_{\Lambda,\epsilon}(f),$$

(2.4)

where

$$E_{\Lambda,\epsilon}(f) = \inf \{\|f - p\|_{L^\infty(\Omega)} + \epsilon\|p\|_{L^2(D,\nu)} : p \in P_{\Lambda}\},$$

and $\mu$ is the measure given by (3.2). Moreover, the coefficients $c^\epsilon$ of $f_{T,\Lambda,\epsilon}$ satisfy

$$\|c^\epsilon\|_2 = \|f_{T,\Lambda,\epsilon}\|_{L^2(D,\nu)} \leq \frac{E_{\Lambda,\epsilon}(f)}{\epsilon},$$

(2.5)

and the absolute ($\ell^2, L^2$)-condition number of the reconstruction operator $L_{T,\Lambda,\epsilon} : \mathbb{C}^M \to P_{\Lambda}, b \mapsto f_{T,\Lambda,\epsilon}$, where $b$ is as in [2.2], is at most $C_{T,\Lambda,\epsilon}$.

See [4] Several remarks are in order. First, the bound (2.4) separates the accuracy of the regularized least-squares approximation into an approximation error term $E_{\Lambda,\epsilon}(f)$ depending only on $\epsilon$ and the space $P_{\Lambda}$ and independent of the samples $\Upsilon$, and a constant $C_{T,\Lambda,\epsilon}$ depending on $\epsilon$, $\Upsilon$ and $P_{\Lambda}$. In other words, $E_{\Lambda,\epsilon}(f)$ determines the rate of approximation, whereas $C_{T,\Lambda,\epsilon}$ (more specifically, the requirement that $C_{T,\Lambda,\epsilon} \lesssim 1$) determines the sample complexity.

Second, notice that the error term $E_{\Lambda,\epsilon}(f)$ depends on how well $f$ can be approximated in $\Omega$ by polynomials $p \in P_{\Lambda}$ that do not grow too large on $D$. The latter requirement – which stems from the regularization carried out – is an expression of stability, since a polynomial growing large on $D$ would necessarily have large coefficients. Our main estimates for $E_{\Lambda,\epsilon}(f)$, given below, are derived by constructing polynomials which approximate $f$ at suitable rates in $\Omega$ (depending on the smoothness of $f$), but remain bounded on $D$.

Finally, we remark that (2.5) ensures the stored values – namely, the coefficients $c^\epsilon$ – cannot be too large in magnitude, which would otherwise result in ill-conditioning of the evaluation map $c^\epsilon \mapsto f_{T,\Lambda,\epsilon}(\mathbf{x})$. While $\|c^\epsilon\|_2$ may be of magnitude roughly $1/\epsilon$ initially, once the approximation error $E_{\Lambda,\epsilon}(f)$ reaches close to the target accuracy $\epsilon$ we have $\|c^\epsilon\|_2 \lesssim 1$.

**Rate of approximation.** In [5] we address the behaviour of the approximation error term $E_{\Lambda,\epsilon}(f)$ for the main example considered in this paper: namely, Legendre polynomials on $D = (-1,1)^d$.

We consider two standard choices of index sets $\Lambda$: the total degree index set $\Lambda = \Lambda_{n}^{TD}$ defined in (3.5) and the hyperbolic cross index set $\Lambda = \Lambda_{n}^{HC}$ defined in (3.6). The former is suitable for low-dimensional problems, but quickly becomes too large as $d$ increases. The cardinality of the latter on the other hand scales much more mildly with $d$.

Our main results are split into two cases:

(i) $f$ smooth in $\Omega$ only. In the first case, $f$ is smooth in $\Omega$ but may be nonsmooth, or even undefined in $D \setminus \Omega$. If $\Omega$ is a Lipschitz domain and $f \in H^m(\Omega,\mu)$, where $H^m(\Omega,\mu)$ is the classical Sobolev...
space of order \( m \) (see (5.1)), then

\[
E_{\Lambda,\epsilon}(f) \leq \begin{cases} 
  c_{m,d,\Omega} \left( n^{d-m} + \epsilon \right) \|f\|_{H^m(\Omega,\nu)} & \Lambda = \Lambda_n^{TD} \\
  c_{m,d,\Omega} \left( n^{d-m} + \epsilon \right) \|f\|_{H^m(\Omega,\nu)} & \Lambda = \Lambda_n^{HC}
\end{cases}, \tag{2.6}
\]

where \( c_{m,d,\Omega} > 0 \) is a constant depending on \( m, d \) and \( \Omega \) but independent of \( f \). See Theorem 5.1 (we note in passing that the factor \( d - m \) can be improved slightly to \( \theta(d) - m \) where \( \theta(d) \) is a particular constant satisfying \( \theta(d) \leq d \)). This result asserts convergence at an algebraic rate depending on the smoothness of \( f \) in \( \Omega \) only. However, it also exhibits the familiar curse of dimensionality. In the case of the total degree index set \( \Lambda^{TD} \) the cardinality \( N = |\Lambda^{TD}| \asymp n^d \) as \( n \to \infty \), and therefore

\[
n^{-m} \asymp N^{-m}, \quad n \to \infty,
\]

whereas for the hyperbolic cross space (wherein \( N = |\Lambda^{HC}_n| \asymp n(\log(n))^{d-1} \)) one has

\[
n^{-m} \asymp N^{-m}(\log(N))^{(m-d)(d-1)/d}, \quad n \to \infty.
\]

\((ii) f \) smooth in \( D \). In high-dimensional approximation a standard way to overcome the \( d \)-dependence in results such as (2.6) is to assume certain anisotropic smoothness. As we discuss in [9] it is currently unknown how to do this within the setting of case \((i)\). However, when \( f \) has appropriate regularity over the whole of \( D \) – or equivalently \( f \) is the restriction to \( \Omega \) of some appropriately regular function defined on \( D \) – then we have the following result. If \( f \in H^m_{\text{mix}}(D,\nu) \), where \( H^m_{\text{mix}}(D,\nu) \) is the Sobolev space of dominating mixed smoothness on \( D \) (see (5.2)), then

\[
E_{\Lambda,\epsilon}(f) \leq \begin{cases} 
  c_{m,d}\|f\|_{H^m_{\text{mix}}(D,\nu)} n^{1-m} + \epsilon\|f\|_{L^2(D,\nu)} & \Lambda = \Lambda_n^{TD} \\
  c_{m,d}\|f\|_{H^m_{\text{mix}}(D,\nu)} n^{1-m}(\log(n))^{\frac{d-1}{2}} + \epsilon\|f\|_{L^2(D,\nu)} & \Lambda = \Lambda_n^{HC}
\end{cases}, \tag{2.7}
\]

where \( c_{m,d} > 0 \) is a constant depending on \( m \) and \( d \) but independent of \( \Omega \) and \( f \). See Theorem 5.3. Note that

\[
n^{-m} \asymp N^{1-m}, \quad N = |\Lambda^{TD}_n|,
\]

whereas

\[
n^{1-m}(\log(n))^{\frac{d-1}{2}} \asymp N^{1-m}(\log(n))^{(d-1)(m-1/2)}, \quad N = |\Lambda^{HC}_n|.
\]

Hence, up to the logarithmic factor, the hyperbolic cross index set \( \Lambda^{HC}_n \) achieves an algebraic rate of convergence that is independent of the dimension \( d \), and therefore suitable for higher-dimensional computations. Our numerical results in [8] show computations using the hyperbolic cross index set for dimensions up to \( d = 15 \).

Let us make several remarks. First, we note that case \((ii)\) requires absolutely no conditions on the domain \( \Omega \), besides being measurable. In particular, the domain can be extremely rough, as long as \( f \) is smooth over the whole extended domain \( D \). In [8] we show some numerical results of this type. Second, both (2.6) and (2.7) exhibit the algebraic factor \( n^{-1-m} \). The additional power of \( n \) stems from the appearance of the \( L^\infty(\Omega) \) norm in \( E_{\Lambda,\epsilon}(f) \). This factor can be improved whenever \( \Omega \) is compactly contained in \( D \), in which case one obtains a factor of the form \( n^{1/2-m} \) (see Theorems 5.1 and 5.3). Third, when the sample points are drawn randomly and independently (as they are in this paper) it is possible to prove an estimate in expectation for the squared \( L^2 \)-error of a related estimator (see (7)) based on the \( L^2 \)-norm approximation error

\[
\tilde{E}_{\Lambda,\epsilon}(f) = \inf \left\{ \|f - p\|_{L^2(\Omega,\mu)} + \epsilon\|p\|_{L^2(D,\nu)} : p \in P_\Lambda \right\}.
\]

6
Figure 1: Examples of domains that have the \( \lambda \)-rectangle property.

See Theorem 7.1. Analogous to (2.6) and (2.7), this approximation error admits the following estimates. First, if \( \Omega \) is Lipschitz and \( f \in H^m(\Omega, \mu) \) then

\[
\tilde{E}_{\Lambda, \epsilon}(f) \leq \begin{cases} 
  c_{m,d,\Omega} (n^{-m} + \epsilon) \|f\|_{H^m(\Omega, \mu)} & \Lambda = \Lambda^n_{TD} \\
  c_{m,d,\Omega} \left( n^{-\frac{m}{d}} + \epsilon \right) \|f\|_{H^m(\Omega, \mu)} & \Lambda = \Lambda^n_{HC} 
\end{cases} \quad (2.8)
\]

Conversely, if \( f \in H^m_{mix}(D, \nu) \) then

\[
\tilde{E}_{\Lambda, \epsilon}(f) \leq c_{m,d} \|f\|_{H^m_{mix}(D, \nu)} n^{m} + \epsilon \|f\|_{L^2(D, \nu)}, \quad \Lambda = \Lambda^n_{TD} \text{ or } \Lambda = \Lambda^n_{HC}. \quad (2.9)
\]

See Theorems 7.2 and 7.3. As with \( E_{\Lambda, \epsilon}(f) \) above, this latter result for the hyperbolic cross index set shows how the polynomial frame approximation can mitigate the curse of dimensionality.

**Sample complexity.** Our final result concerns efficiency, i.e. sample complexity, of the approximation. In view of Theorem 2.2, this corresponds to estimating how many samples are required in order for the condition \( C_{Y, \Lambda, \epsilon} \lesssim 1 \) to hold. Our main results in this direction are for the following class of domains \( \Omega \):

**Definition 2.3 (\( \lambda \)-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

\[
\inf_{R \in \mathcal{R}} \text{Vol}(R) = \lambda \text{Vol}(\Omega).
\]

Note that many domains of practical interest have this property (see Fig. 1 for several examples), however there are some notable exceptions, including simplices and balls. See §6.3 for further discussion. As we show in §6.2 when the samples \( y_m \) are chosen randomly and independently according to the uniform measure on \( \Omega \) the sample complexity of the approximation can in general be related to the constant of the \( L^\infty - L^2 \) Nikolskii-type inequality for the space \( P_\Lambda \) – a fact that is well-known for tensor-product domains (see, for example, [21]) but which we extend to irregular domains. We use the \( \lambda \)-rectangle property to get concrete estimates for this constant, culminating in the following result:

**Theorem 2.4 (Sample complexity).** Suppose that \( \Omega \subseteq (-1, 1)^d \) has the \( \lambda \)-rectangle property and let \( P_\Lambda \) be constructed from the tensor Legendre polynomial basis on \((-1, 1)^d\), where \( \Lambda \subset \mathbb{N}_0^d \) is any lower set (see Definition 3.1) of cardinality \( |\Lambda| = N \). Let \( 0 < \delta, \gamma < 1 \) and \( y_1, \ldots, y_M \) be independent and randomly drawn according to the uniform probability measure on \( \Omega \). Then

\[
C_{Y, \Lambda, \epsilon} \leq \frac{1}{\sqrt{1 - \delta}}, \quad \forall \epsilon > 0,
\]
with probability at least $1 - \gamma$, provided

$$M \geq N^{2\lambda - 1} \left((1 - \delta) \log(1 - \delta) + \delta \right)^{-1} \log(N/\gamma).$$

See Corollary 6.7. This result establishes quadratic scaling of the number of samples with the dimension of the polynomial space. Note that this result holds for all lower sets, and in particular, the total degree and hyperbolic cross index sets discussed above.

2.4 Related work

The idea of approximating a function on an irregular domain by using an orthogonal basis on a tensor-product domain was considered within the context of embedded or fictitious domain methods in numerical PDEs [36] (see also [11]). So-called Fourier extensions or Fourier continuations were studied in detail in [12, 14]. Applications to surface parametrization and numerical PDEs in complex geometries were considered in [14] and [10, 13, 30] respectively. Our work can be considered an extension of [7] from the univariate to the multivariate setting, although we use algebraic as opposed to trigonometric polynomials since these are more common in applications such as UQ.

Our work also extends recent research on computing polynomial approximations of functions defined on high-dimensional tensor-product domains. This approach has received substantial interest recently, due to its applications in, notably, UQ. See [11, 17, 22, 21, 26, 35, 47] and references therein. One particular consequence of our work is to show that an irregular domain (either known or unknown) is no barrier to polynomial approximation of high-dimensional functions. Note that polynomial approximations are frequently used in practical UQ studies even when the domain is non-tensorial (see [39, 40] and references therein). Our work therefore provides a theoretical basis for these approaches.

Finally, we note that the interpretation of polynomial frame approximations are just once instance of so-called numerical frame approximations. For further details, including other uses of frames in numerical analysis and approximation, we refer to [5, 6].

3 Polynomial frame approximations

3.1 Notation

We first require some further notation. Throughout this paper $D \subseteq \mathbb{R}^d$ will be a domain with a probability measure $\nu$. Typically, $D$ will be of tensor-product type, i.e.

$$D = [a_1, b_1] \otimes \cdots \otimes [a_d, b_d] \subseteq \mathbb{R}^d$$

where $-\infty \leq a_k < b_k \leq \infty$ and $\nu = \nu^{(1)} \otimes \cdots \otimes \nu^{(d)}$ will be a tensor-product of one-dimensional probability measures. We write $L^2(D, \nu)$ for the space of square-integrable functions on $D$.

The $d$-dimensional variable is denoted by $y = (y_1, \ldots, y_d) \in \mathbb{R}^d$. Given $D$, we let $\Omega \subseteq D$ be a domain and define the probability measure $\mu$ by

$$d\mu(y) = \frac{\mathbb{I}_\Omega(y)}{v_\Omega} \, d\nu(y), \quad v_\Omega = \int_\Omega d\nu,$$  

where $\mathbb{I}_\Omega$ is the indicator function of $\Omega$. We write $L^2(D, \mu)$ for the space of square-integrable functions on $D$. 

8
Throughout, \( \mathbf{n} = (n_1, \ldots, n_d) \in \mathbb{N}_0^d \) denotes a multi-index. Let \( I \subseteq \mathbb{N}_0^d \) be a countable set of multi-indices and \( \{ \psi_n \}_{n \in I} \) be an orthonormal basis of \( L^2(D, \nu) \). If \( D \) is of the form \( (3.1) \), then this basis will usually be of tensor product-type, i.e.

\[
\psi_n(y) = \prod_{k=1}^d \psi_{n_k}^{(k)}(y_k),
\]

where \( \{ \psi_{n_k}^{(k)} \} \) is an orthonormal basis of \( L^2((a_k, b_k), \nu^{(k)}) \). Given \( \{ \psi_n \}_{n \in I} \) we let

\[
\phi_n = \psi_n|_{\Omega}, \quad n \in I,
\]

be the corresponding functions defined on \( \Omega \).

### 3.2 Multi-index sets

Our interest lies in computing finite approximations from the system \( (3.3) \). To this end, let \( \Lambda \subset I \) be a finite multi-index set and define

\[
P_{\Lambda} = \text{span}\{ \phi_n : n \in \Lambda \} \subset L^2(\Omega, \mu),
\]

as the finite-dimensional space within which we seek an approximation to \( f \). We shall consider the following three standard choices of multi-index sets. The tensor product set

\[
\Lambda = \Lambda_n^{\text{TP}} = \left\{ n \in \mathbb{N}_0^d : |n|_{\infty} \leq n \right\},
\]

where \( |n|_{\infty} = \max_{k=1, \ldots, d} |n_k| \), the total degree set

\[
\Lambda = \Lambda_n^{\text{TD}} = \left\{ n \in \mathbb{N}_0^d : |n|_1 \leq n \right\},
\]

where \( |n|_1 = |n_1| + \ldots + |n_d| \), and the (isotropic) hyperbolic cross set

\[
\Lambda = \Lambda_n^{\text{HC}} = \left\{ n \in \mathbb{N}_0^d : |n|_{\text{hc}} \leq n + 1 \right\}, \quad |n|_{\text{hc}} = \prod_{k=1}^d (|n_k| + 1).
\]

Note that the cardinality \( N = |\Lambda_n^{\text{TP}}| = (n+1)^d \) usually grows too quickly with \( n \) in high dimensions to be practical. The total degree set, with cardinality

\[
N = |\Lambda_n^{\text{TD}}| = \binom{n+d}{d},
\]

mitigates this issue to some extent, but still typically grows too rapidly for moderate to high-dimensional problems. Hyperbolic cross index sets are a practical alternative in this case. An exact formula for the cardinality of the hyperbolic cross \( \Lambda_n^{\text{HC}} \) in terms of \( n \) and \( d \) is not known, but there are a variety of upper bounds, including:

\[
|\Lambda_n^{\text{HC}}| \leq \left[ (n + 1)(1 + \log(n + 1))^{d-1} \right].
\]

See, for example, [31, Prop. A.1].

The above three multi-index sets are all examples of so-called \textit{lower sets} (also known as \textit{downward closed} or \textit{monotone} sets – see, for example, [22, 27]):
Since there are infinitely many extensions of \( f \) in the frame \( \{ \psi_n \}_{n \in I} \), making this system a frame. Frames such as this for which \( A = B \) are known as tight frames.

A general property of frames is their redundancy: any \( f \in H \) can have infinitely many expansions \( f = \sum_{n \in I} c_n \phi_n \) with coefficients \( \{ c_n \}_{n \in I} \in l^2(I) \). It is straightforward to see why redundancy occurs in the polynomial frame. Indeed, let \( \tilde{f} \) be any extension of \( f \) to \( L^2(D, \nu) \) and define

\[
c_n = \langle \tilde{f}, \psi_n \rangle_{L^2(D, \nu)},
\]

as the coefficients of \( \tilde{f} \) in the orthonormal basis \( \{ \psi_n \}_{n \in I} \). Then

\[
\sum_{n \in I} c_n \phi_n = \sum_{n \in I} c_n \psi_n \big|_\Omega = \tilde{f} |_\Omega = f.
\]

Since there are infinitely many extensions of \( f \) to \( L^2(D, \nu) \), each with distinct coefficients \( \{ c_n \}_{n \in I} \), it follows that there are infinitely many representations of \( f \) in the frame \( \{ \phi_n \}_{n \in I} \).

### 3.4 Least-squares polynomial frame approximations

Let \( \mathcal{Y} = \{ y_1, \ldots, y_M \} \subset \Omega \) be a set of \( M \) distinct points (for the moment we choose not to specify their distribution) and \( \Lambda \) be a finite set of multi-indices of size \(|\Lambda| = N\), where \( N \leq M \). We compute an approximation to \( f \) in the space \( P_\Lambda \) by discrete least-squares fitting:

\[
f_{\mathcal{Y}, \Lambda} = \arg\min_{p \in P_\Lambda} \frac{1}{M} \sum_{y \in \mathcal{Y}} |f(y) - p(y)|^2.
\]
If \( f_{\Upsilon,\Lambda} \) is expressed as
\[
f_{\Upsilon,\Lambda} = \sum_{n \in \Lambda} c_n \phi_n,
\]
then this is equivalent to the algebraic least-squares problem
\[
c = (c_n)_{n \in \Lambda} = \arg\min_{x \in \mathbb{C}^N} \|Ax - b\|_2,
\]
where
\[
A = A_{\Upsilon,\Lambda} = \left( \frac{1}{\sqrt{M}} \phi_n(y) \right)_{y \in \Upsilon, n \in \Lambda} \in \mathbb{C}^{M \times N}, \quad b = b_{\Upsilon} = \left( \frac{1}{\sqrt{M}} f(y) \right)_{y \in \Upsilon} \in \mathbb{C}^M.
\]
As mentioned in \( \S 2.2 \) and shown explicitly in \( \S 4.1 \) below, the matrix \( A \) is typically severely ill-conditioned for large \( N \). Hence it is necessary to regularize \( (3.8) \). We shall do this via truncated singular value decompositions (i.e. spectral filtering).

To this end, suppose that \( A \) have singular values \( \{\sigma_n\}_{n \in \Lambda} \) and singular value decomposition \( A = U \Sigma V^* \), where \( U \in \mathbb{C}^{M \times M} \), \( \Sigma \in \mathbb{R}^{M \times N} \) and \( V \in \mathbb{C}^{N \times N} \). Define
\[
A^\epsilon_{\Upsilon,\Lambda} = U \Sigma^\epsilon V^*,
\]
where the diagonal matrix \( \Sigma^\epsilon \) has \( n \)-th entry \( \sigma_n \) if \( \sigma_n > \epsilon \) and zero otherwise. Then the coefficients \( c_\epsilon \) of truncated SVD least-squares approximation are given by
\[
c_\epsilon = (A^\epsilon_{\Upsilon,\Lambda})^+ b_{\Upsilon} = V (\Sigma^\epsilon)^+ U^* b_{\Upsilon},
\]
where \( \dag \) denotes the pseudoinverse. Correspondingly, we define the truncated SVD least-squares approximation as
\[
f_{\Upsilon,\Lambda,\epsilon} = \sum_{n \in \Lambda} (c_\epsilon)_n \phi_n.
\]  

We consider this approximation from now on. Note that the regularization parameter \( \epsilon \) is usually set in relation to some desired target accuracy (see \( \S 8 \)).

**Remark 3.3** Related strategies such as Tikhonov regularization could be used instead, with only minor modifications to the ensuing presentation.

### 3.5 Main example

We end this section by introducing the example we use to illustrate our main results. This is the case where \( \Omega \) is bounded and, without loss of generality, contained in \( D = (-1,1)^d \), and where \( \{\psi_n\}_{n \in \mathbb{N}_0^d} \) is the tensor Legendre polynomial basis on \( D \). When normalized with respect to the uniform probability measure on \( D \), this is defined by
\[
\psi_n(y) = \prod_{k=1}^d \sqrt{2n_k + 1} P_{n_k}(y_k),
\]
where \( P_n \) is the \( n \)-th classical Legendre polynomial. For the truncated index set, we let \( \Lambda = \Lambda_{n}^{TD} \) or \( \Lambda = \Lambda_{n}^{HC} \) be either the total degree (3.5) or hyperbolic cross (3.6) index set with index \( n \). We also assume that the sampling points \( y_1, \ldots, y_M \) are drawn independently according to the uniform probability measure on \( \Omega \):
\[
d\mu(y) = \frac{1}{\text{Vol}(\Omega)} \, dy.
\]
While this approach leads to concrete, \( d \)-independent sample complexity estimates for many domains, we do not claim that this is optimal. See \( \S 9 \) for further discussion.
4 Accuracy and conditioning

We now address the accuracy and conditioning of the approximation (3.9). In §4.1 we first show that least-squares matrix $A$ is ill-conditioned for large $N$, thus explaining why regularization is needed. Next, in §4.2 we introduce a constant $C_{T,A,\epsilon}$ which controls conditioning and accuracy. We then present our main result in §4.3. Note that our approach in §4.2 and §4.3 follows that of [6] (which applies to general frames).

4.1 Ill-conditioning of the matrix $A$

Frame approximations always lead to ill-conditioned least-squares matrices [5]. In the case of the polynomial frame, this is related to the Remez inequality for the polynomial space $P_N$ over $\Omega$ and $D$. To see this, observe that the minimal and maximal singular values of $A$ are

$$
\sigma_{\min}(A) = \inf_{p \in P_N \atop p \neq 0} \left\{ \sqrt{\frac{1}{M} \sum_{y \in \mathcal{T}} |p(y)|^2} \right\}, \quad \sigma_{\max}(A) = \sup_{p \in P_N \atop p \neq 0} \left\{ \sqrt{\frac{1}{M} \sum_{y \in \mathcal{T}} |p(y)|^2} \right\}.
$$

For simplicity, assume that the constant function is contained in $P_N$. This will hold in all examples considered later. Letting $p(y) = 1$ we get $\sigma_{\max}(A) \geq 1$. Conversely, note that $\frac{1}{M} \sum_{y \in \mathcal{T}} |p(y)|^2 \leq \|p\|_{L^\infty(\Omega)}^2$ and let $N(P_N, D, \nu) > 0$ be a constant depending on $P_N$, $D$ and $\nu$ such that

$$
\|p\|_{L^\infty(D)} \leq N(P_N, D, \nu) \|p\|_{L^2(D, \nu)}, \quad \forall p \in P_N.
$$

We refer to this as a Nikolskii-type inequality. Inequalities such as these will be discussed further in §6 since they are pivotal in estimating the sample complexity of the approximation. This gives

$$
\frac{1}{\sigma_{\min}(A)} \geq (N(P_N, D, \nu))^{-1} \sup \left\{ \frac{\|p\|_{L^\infty(D)}}{\|p\|_{L^\infty(\Omega)}} : p \in P_N, \ p \neq 0 \right\},
$$

and therefore the condition number of $A$ satisfies

$$
\text{cond}(A) \geq \frac{R(P_N, \Omega, D)}{N(P_N, D, \nu)}, \quad (4.1)
$$

where $R(P_N, \Omega, D)$ is the constant in Remez’s inequality for the domains $\Omega$ and $D$:

$$
\|p\|_{L^\infty(D)} \leq R(P_N, \Omega, D) \|p\|_{L^\infty(\Omega)}, \quad p \in P_N.
$$

Note that the bound (4.1) is deterministic, and independent of the number of samples $M$.

Typically, the right-hand side of (4.1) will grow rapidly with $N$. To see why, note first that the constant in the Nikolskii-type inequality is usually expected to be at most algebraic in $N = |\Lambda|$. In particular, if $\nu$ is the uniform measure on $D$ and $\Lambda$ is a lower set, then $N(P_N, D, \nu) \leq N^2$ [32, Thm. 6] (see also the proof of Theorem 6.6). Similar bounds are found in [32] for other ultraspherical and Jacobi measures. Conversely, the constant $R(P_N, \Omega, D)$ in Remez’s inequality is typically exponentially-large in $N$. The exact nature of this behaviour depends on the domain $\Omega$ and the index set $\Lambda$, and for the sake of brevity, we will not give a more detailed discussion here. However, we note in passing that in the one-dimensional case for example, if $\Lambda = \{0, \ldots, N - 1\}$ and $D = (-1, 1)$ then

$$
R(P_N, \Omega, (-1, 1)) \leq T_N(4/|\Omega| - 1),
$$

12
where \( T_{N-1} \) is the \((N - 1)\)th Chebyshev polynomial and \( |\Omega| \) denotes the Lebesgue measure of \( \Omega \). Moreover, equality holds if \( \Omega = [-1, -1 + |\Omega|] \) and in which case one has the exponential growth

\[
R(P_\Lambda, \Omega, (-1, 1)) \geq \frac{1}{2} \left( \frac{4}{|\Omega|} - 1 \right)^{N-1}.
\]

We refer to [28] for further information, including results in higher dimensions, as well as to [42] for results on multivariate Remez inequalities for hyperbolic cross index sets.

### 4.2 Key constants

For convenience we now define the following operator

\[
T_\Lambda : \mathbb{C}^N \to P_\Lambda, \ c = \{c_n\}_{n \in \Lambda} \mapsto \sum_{n \in \Lambda} c_n \phi_n.
\]

This is commonly referred to as the \textit{synthesis} operator in frame theory. We now let

\[
C_{T, \Lambda, \epsilon} = \max \{ C'_{T, \Lambda, \epsilon}, C''_{T, \Lambda, \epsilon} \},
\]

where

\[
C'_{T, \Lambda, \epsilon} = \max_{b \in \mathbb{C}^M, \|b\|_2 = 1} \| T_\Lambda(A_{T, \Lambda, \epsilon})^\dagger b \|_{L^2(\Omega, \mu)}
\]

\[
C''_{T, \Lambda, \epsilon} = \epsilon^{-1} \max_{d \in \mathbb{C}^N, \|d\|_2 = 1} \| T_\Lambda d - T_\Lambda(A_{T, \Lambda, \epsilon})^\dagger A_{T, \Lambda} d \|_{L^2(\Omega, \mu)}.
\]

It is useful to reinterpret these constants. First, define the \textit{reconstruction operator}

\[
\mathcal{L}_{T, \Lambda, \epsilon} : \mathbb{C}^M \to P_\Lambda; \ b \mapsto T_\Lambda(A_{T, \Lambda, \epsilon})^\dagger b.
\]

This operator takes a vector of samples \( b \in \mathbb{C}^M \) to its truncated SVD approximation in \( P_\Lambda \). In particular, if

\[
S_T : L^\infty(\Omega) \to \mathbb{C}^M; \ f \mapsto \left\{ \frac{1}{\sqrt{M}} f(y) \right\}_{y \in \Upsilon},
\]

is the operator taking a function \( f \) to its samples then

\[
f_{T, \Lambda, \epsilon} = \mathcal{L}_{T, \Lambda, \epsilon} S_T f.
\]

Note that \( C'_{T, \Lambda, \epsilon} \) is precisely the operator norm – or equivalently, since it is a linear operator, the absolute condition number – of \( \mathcal{L}_{T, \Lambda, \epsilon} \) with resect to the \( \ell^2 \) and \( L^2(\Omega, \mu) \)-norms:

\[
C'_{T, \Lambda, \epsilon} = \max_{b \in \mathbb{C}^M, \|b\|_2 = 1} \| \mathcal{L}_{T, \Lambda, \epsilon} b \|_{L^2(\Omega, \mu)}.
\]

In other words, boundedness of \( C_{T, \Lambda, \epsilon} \) implies robustness of the approximation to perturbations in the data (e.g. noise). On the other hand, \( C''_{T, \Lambda, \epsilon} \) also has the equivalent definition

\[
C''_{T, \Lambda, \epsilon} = \epsilon^{-1} \sup \left\{ \| p - p_{T, \Lambda, \epsilon} \|_{L^2(\Omega, \mu)} : p \in P_\Lambda, \| p \|_{L^2(D, \nu)} = 1 \right\}.
\]

In particular, \( C''_{T, \Lambda, 0} = 0 \) since the unregularized mapping \( f \mapsto f_{T, \Lambda, 0} \) is a projection onto \( P_\Lambda \). For \( \epsilon > 0 \) this constant measures how close the mapping \( f \mapsto f_{T, \Lambda, \epsilon} \) is to being a projection onto \( P_\Lambda \).
4.3 Main result on accuracy and conditioning

**Theorem 4.1.** Let \( f \in L^{\infty}(\Omega, \mu) \) and suppose that \( f_{T, \Lambda, \epsilon} \) is the truncated SVD least-squares approximation. Then

\[
\|f - f_{T, \Lambda, \epsilon}\|_{L^2(\Omega, \mu)} \leq (1 + C_{T, \Lambda, \epsilon}') \|f - p\|_{L^\infty(\Omega)} + \epsilon C_{T, \Lambda, \epsilon}'' |p|_{L^2(D, \nu)}
\]

\[
\leq (1 + C_{T, \Lambda, \epsilon}) E_{\Lambda, \epsilon}(f),
\]

where \( C_{T, \Lambda, \epsilon}', C_{T, \Lambda, \epsilon}'' \) and \( C_{T, \Lambda, \epsilon} \) are as in (4.3) and (4.2) respectively, and

\[
E_{\Lambda, \epsilon}(f) = \inf \left\{ \|f - p\|_{L^\infty(\Omega)} + \epsilon |p|_{L^2(D, \nu)} : p \in P_{\Lambda} \right\}.
\]

Moreover, the coefficients \( c^\epsilon \) of \( f_{T, \Lambda, \epsilon} \) satisfy

\[
\|c^\epsilon\|_2 = \|f_{T, \Lambda, \epsilon}\|_{L^2(D, \nu)} \leq \frac{E_{\Lambda, \epsilon}(f)}{\epsilon}.
\]

**Proof.** Let \( p = T_{\Lambda} e \) for some \( e \in \mathbb{C}^N \). Then, recalling the definitions of the constants \( C_{T, \Lambda, \epsilon}' \) and \( C_{T, \Lambda, \epsilon}'' \), we have

\[
\|f - f_{T, \Lambda, \epsilon}\|_{L^2(\Omega, \mu)} \leq \|f - p\|_{L^2(\Omega, \mu)} + \|p_{T, \Lambda, \epsilon} - f_{T, \Lambda, \epsilon}\|_{L^2(\Omega, \mu)} + \|p - p_{T, \Lambda, \epsilon}\|_{L^2(\Omega, \mu)}
\]

\[
\leq \|f - p\|_{L^2(\Omega, \mu)} + C_{T, \Lambda, \epsilon}' \|S_{\Lambda}(f - p)\|_2 + \epsilon C_{T, \Lambda, \epsilon}'' |p|_{L^2(D, \nu)}
\]

\[
\leq (1 + C_{T, \Lambda, \epsilon}) \|f - p\|_{L^\infty(\Omega)} + \epsilon C_{T, \Lambda, \epsilon}'' |p|_{L^2(D, \nu)},
\]

which gives the first result. Note that in the third step we use (4.5) to deduce that \( \|S_{\Lambda}(f - p)\|_2 \leq \|f - p\|_{L^\infty(\Omega)} \) and the fact that \( \mu \) is a probability measure, which implies that \( \|f - p\|_{L^2(\Omega, \mu)} \leq \|f - p\|_{L^\infty(\Omega)} \). For the second result, we first use Parseval’s identity to give \( \|c^\epsilon\|_2 = \|f_{T, \Lambda, \epsilon}\|_{L^2(D, \nu)} \)

and then write

\[
\|f_{T, \Lambda, \epsilon}\|_{L^2(D, \nu)} \leq \|f_{T, \Lambda, \epsilon} - p_{T, \Lambda, \epsilon}\|_{L^2(D, \nu)} + \|p_{T, \Lambda, \epsilon}\|_{L^2(D, \nu)}.
\]

Consider the first term. By (4.6) we have

\[
\|f_{T, \Lambda, \epsilon} - p_{T, \Lambda, \epsilon}\|_{L^2(D, \nu)} = \left\| T_{\Lambda} (A_{T, \Lambda, \epsilon})^\dagger S_{\Lambda}(f - p) \right\|_{L^2(D, \nu)}
\]

\[
= \left\| (A_{T, \Lambda, \epsilon})^\dagger S_{\Lambda}(f - p) \right\|_2 \leq \frac{1}{\epsilon} \|S_{\Lambda}(f - p)\|_2 \leq \frac{1}{\epsilon} \|f - p\|_{L^\infty(\Omega)}.
\]

Here in the second step we use Parseval’s identity, in the third step we use standard properties of the SVD and in the fourth step we use (4.5). Now consider the second term of (4.8). Observe that \( A_{T, \Lambda} = S_{\Lambda} T_{\Lambda} \). Hence, if \( p = T_{\Lambda} c \) then, using standard properties of the SVD once more, we get

\[
\|p_{T, \Lambda, \epsilon}\|_{L^2(D, \nu)} = \left\| T_{\Lambda} (A_{T, \Lambda, \epsilon})^\dagger S_{\Lambda} T_{\Lambda} c \right\|_{L^2(D, \nu)} = \left\| (A_{T, \Lambda, \epsilon})^\dagger A_{T, \Lambda} c \right\|_2 \leq \|c\|_2 = \|p\|_{L^2(D, \nu)}.
\]

Combining this with (4.9) and substituting both into (4.8) now gives the second result. \(\square\)

A few remarks are in order. First, to ensure accuracy and good (absolute) conditioning of the approximation we need to ensure that \( C_{T, \Lambda, \epsilon} \leq 1 \). This constant depends on the polynomial space, the data and the threshold \( \epsilon \), but is independent of the function \( f \). In §5 we derive bounds for this constant. Second, once \( C_{T, \Lambda, \epsilon} \) is bounded, the approximation error is determined via the term \( E_{\Lambda, \epsilon}(f) \), which depends on \( f \) and the polynomial space but is independent of the data. We estimate this term for functions in certain Sobolev spaces in §5.
Third, we notice that coefficients of the ensuing approximation are bounded by the approximation error divided by $\epsilon$. Thus, although the coefficients may initially be $O(1/\epsilon)$, they are $O(1)$ in the limit as the dimension $N$ of the approximation space $P_\Lambda$ tends to infinity. Note that bounded coefficients are particularly important for practical computations, since these are the values that will be stored. Indeed, if the coefficients could grow arbitrarily large in relation to the function $f$ then the pointwise evaluation operator $\epsilon^\infty \mapsto f_{\infty,\Lambda,\epsilon}(x)$ would be ill-conditioned.

Fourth and finally, we note in passing that $C_{\infty,\Lambda,\epsilon} \leq 1/(v_\Omega \epsilon)$ for any $\Upsilon, \Lambda$ and $\epsilon > 0$. In other words, the ill-conditioning of the reconstruction operator is at worst $1/\epsilon$.

5 Approximation error for Legendre polynomial frames

We now consider the approximation error $E_{\Lambda,\epsilon}(f)$, defined by (4.7). In doing so, we shall treat the following two scenarios separately:

(i) $f$ defined and smooth over $D$,

(ii) $f$ undefined or nonsmooth over $D$.

We first require several notions of smoothness. Let

$$H^m(\Omega, \mu) = \{ f \in L^2(\Omega, \mu) : D^j f \in L^2(\Omega, \mu) : |j|_1 \leq m \},$$

be the classical Sobolev spaces of index $m \geq 0$ on $\Omega$, with norm

$$\|f\|_{H^m(\Omega, \mu)} = \sqrt{\sum_{|j|_1 \leq m} \|D^j f\|^2_{L^2(\Omega, \mu)}}.$$

Here $D^j = \frac{\partial^{|j|_1}}{\partial x_1^{j_1} \cdots \partial x_d^{j_d}}$ is the partial derivative operator of order $j$. These spaces are suitable for approximations using the tensor product or total degree spaces in low dimensions. For moderate to high dimensions, we instead consider Sobolev spaces of dominating mixed smoothness:

$$H^m_{\text{mix}}(\Omega, \mu) = \{ f \in L^2(\Omega, \mu) : D^j f \in L^2(\Omega, \mu) : |j|_\infty \leq m \},$$

with norm

$$\|f\|_{H^m_{\text{mix}}(\Omega, \mu)} = \sqrt{\sum_{|j|_\infty \leq m} \|D^j f\|^2_{L^2(\Omega, \mu)}}.$$

5.1 Results for the classical Sobolev spaces $H^m$

We first consider the tensor product and total degree index sets:

**Theorem 5.1.** Let $P_\Lambda$ be constructed from the tensor Legendre polynomial basis on $L^2(D, \nu)$, where $D = (-1,1)^d$, $\nu$ is the uniform measure on $D$, and $\Lambda = \Lambda_n$ is either the tensor product (3.4) or total degree (3.5) index set of degree $n$. If $\Omega \subseteq D$ and $f \in H^m(D, \nu)$ for some $m > d/2$, then

$$E_{\Lambda,\epsilon}(f) \leq c_{m,d}\|f\|_{H^m(D, \nu)} n^{\theta(d)-m} + \epsilon\|f\|_{L^2(D, \nu)},$$

where

$$\theta(d) = \begin{cases} \frac{d(2d+1)}{2d+2} & \text{odd } d \\ \frac{d(2d+3)}{2d+4} & \text{even } d \end{cases}.$$
Conversely, if $\Omega \subseteq (-1,1)^d$ is Lipschitz and $f \in H^m(\Omega, \mu)$ where $\mu$ is the uniform measure on $\Omega$ and $m > d/2$, then
\[
E_{\Lambda, \epsilon}(f) \leq c_{m,d,\Omega} \left( n^{\theta(d)-m} + \epsilon \right) \|f\|_{H^m(\Omega, \mu)}.
\]

Proof. Let $\Lambda = \Lambda^T_n$. In the first case, since $f$ is defined over the whole of $D$, we may let $p = f_{\Lambda}$ be its orthogonal projection onto $\text{span}\{\psi_n : n \in \Lambda\} \subset L^2(D, \nu)$. Then
\[
E_{\Lambda, \epsilon}(f) \leq \|f - f_{\Lambda}\|_{L^\infty(D)} + \epsilon \|f\|_{L^2(D, \nu)} \leq \|f - f_{\Lambda}\|_{L^\infty(D)} + \epsilon \|f\|_{L^2(D, \nu)}.
\]
(5.4)
It remains to estimate the first term. For this, we first use the Gagliardo–Nirenberg inequality (see, for example, [29]) to give
\[
\|f - f_{\Lambda}\|_{L^\infty(D)} \leq c_{k,d} \|f - f_{\Lambda}\|_{H^k(D, \nu)}^{\frac{d}{k}} \|f - f_{\Lambda}\|_{L^2(D, \nu)}^{1 - \frac{d}{k}}, \quad d < 2k \leq 2m.
\]
We now use the following estimate
\[
\|f - f_{\Lambda}\|_{H^l(D, \nu)} \leq c_{l,m,d} n^{\sigma(l) - m} \|f\|_{H^m(D, \nu)},
\]
where $\sigma(l) = 0$ for $l = 0$ and $\sigma(l) = 2l - 1/2$ for $l > 0$ (see, for example, [16] (5.8.11)). Hence
\[
\|f - f_{\Lambda}\|_{L^\infty(D)} \leq c_{k,m,d} n^{\frac{d(2k-1/2-m)}{2k} - m(1 - \frac{d}{2k})} \|f\|_{H^m(D, \nu)} = c_{k,m,d} n^{\sigma(1 - \frac{d}{2}) - m} \|f\|_{H^m(D, \nu)}.
\]
Setting $k = \frac{d+1}{2}$ (odd $d$) or $k = \frac{d+2}{2}$ (even $d$) and substituting into (5.4) now gives the first result for $\Lambda = \Lambda^T_n$. For the total degree index set $\Lambda = \Lambda^T_{n,d}$ we first recall that $\Lambda^T_{n,d} \subseteq \Lambda^T_n$. We therefore let $p = f_{\Lambda^T_{n,d}}$ so that
\[
E_{\Lambda^T_{n,d}, \epsilon}(f) \leq \|f - f_{\Lambda^T_{n,d}}\|_{L^\infty(D)} + \epsilon \|f\|_{L^2(D, \nu)}.
\]
The result for this index set now follows from the previous bound for $\Lambda^T_{n,d}$.

Now consider the case where $\Omega$ is Lipschitz and $f \in H^m(\Omega, \mu)$. We follow the argument of [5, Prop. 5.8]. We first note that there is an extension $g$ of $f$ to $H^m(D, \nu)$ satisfying
\[
\|g\|_{H^m(D, \nu)} \leq c_{m,d,\Omega} \|f\|_{H^m(\Omega, \mu)}.
\]
Now let $p = g_{\Lambda}$ be the orthogonal projection of $g$ onto $\text{span}\{\psi_n : n \in \Lambda\}$. Then
\[
\|p\|_{L^2(D, \nu)} \leq \|g\|_{L^2(D, \nu)} \leq c_{m,d,\Omega} \|f\|_{H^m(\Omega, \mu)},
\]
and, by the previously-derived result,
\[
\|f - p\|_{L^\infty(\Omega)} \leq \|g - p\|_{L^\infty(\Omega)} \leq c_{m,d,\Omega} n^{\theta(d)-m} \|g\|_{H^m(\Omega, \nu)} \leq c_{m,d,\Omega} n^{\theta(d)-m} \|f\|_{H^m(\Omega, \nu)}.
\]
This gives the second result.

Unsurprisingly, in scenario (i) one obtains a slightly better error bound, where the constant in the $\epsilon$ term involves the smaller $L^2$-norm instead of the $H^m$-norm. For completeness, we now also consider the hyperbolic cross index set:

**Theorem 5.2.** Let $P_n$ be constructed from the tensor Legendre polynomial basis on $L^2(D, \nu)$, where $D = (-1,1)^d$, $\nu$ is the uniform measure on $D$, and $\Lambda = \Lambda_n$ is the hyperbolic cross index set (3.6) of degree $n$. If $\Omega \subseteq D$ and $f \in H^m(D, \nu)$ for some $m > d/2$, then
\[
E_{\Lambda, \epsilon}(f) \leq c_{m,d,\Omega} \|f\|_{H^{m,d}(D, \nu)} n^{\theta(d)-m} \|f\|_{L^2(D, \nu)} + \epsilon \|f\|_{L^2(D, \nu)},
\]
where $\theta(d)$ is as in (5.3). Conversely, if $\Omega \subseteq (-1,1)^d$ is Lipschitz and $f \in H^m(\Omega , \mu)$, where $\mu$ is the uniform measure on $\Omega$ and $m > d/2$, then
\[
E_{\Lambda, \epsilon}(f) \leq c_{m,d,\Omega} \left(n^{\theta(d)-m} + \epsilon \right) \|f\|_{H^m(\Omega, \mu)}.
\]

16
Proof. Let \( n^* = [(n + 1)^{1/d} - 1] \) and observe that \( \Lambda_n^{\text{TP}} \subseteq \Lambda_n^{\text{HC}} \). We now use the arguments from the proof of the previous theorem.

As is to be expected, Theorems 5.1 and 5.2 which assume only classical Sobolev regularity, all exhibit the curse of dimensionality. This can be seen by noting that

\[
n^{\theta(d) - m} \asymp N^{d-m\frac{d}{d-1}}, \quad n \to \infty,
\]

for fixed \( d \) whenever \( \Lambda \) is the total degree or tensor product index set, since in both cases \( N = |\Lambda| \asymp n^d \). Conversely, for the hyperbolic cross index set one has

\[
n^{d-m_{\text{mix}}} \asymp N^{d-m\frac{(d-m)(d-1)}{d}},
\]

since in this case \( N \asymp n(\log(n))^{d-1} \).

5.2 Results for the mixed Sobolev spaces \( H^m_{\text{mix}} \)

Seeking to mitigate the curse of dimensionality when using the hyperbolic cross index set, we now consider the mixed Sobolev spaces \( H^m_{\text{mix}}(\Omega, \mu) \):

**Theorem 5.3.** Let \( P_{\Lambda} \) be constructed from the tensor Legendre polynomial basis on \( L^2(D, \nu) \), where \( D = (-1, 1)^d \) and \( \nu \) is the uniform measure on \( D \). If \( f \in H^m_{\text{mix}}(D, \nu) \) for some \( m \geq 1 \) then

\[
E_{\Lambda, \epsilon}(f) \leq \begin{cases} 
\lambda_{m,d} \| f \|_{H^m_{\text{mix}}(D, \nu)} n^{1-m} + \epsilon \| f \|_{L^2(D, \nu)} & \Lambda = \Lambda_n^{\text{TP}} \text{ or } \Lambda = \Lambda_n^{\text{TD}} \\
\lambda_{m,d} \| f \|_{H^m_{\text{mix}}(D, \nu)} n^{1-m} (\log n)^{\frac{d-1}{d}} + \epsilon \| f \|_{L^2(D, \nu)} & \Lambda = \Lambda_n^{\text{HC}}
\end{cases}
\]

Furthermore, if \( \Omega \) is compactly contained in \( D \), then

\[
E_{\Lambda, \epsilon}(f) \leq \begin{cases} 
\lambda_{m,d,\Omega} \| f \|_{H^m_{\text{mix}}(\Omega, \nu)} n^{1/2-m} + \epsilon \| f \|_{L^2(D, \nu)} & \Lambda = \Lambda_n^{\text{TP}} \text{ or } \Lambda = \Lambda_n^{\text{TD}} \\
\lambda_{m,d,\Omega} \| f \|_{H^m_{\text{mix}}(\Omega, \nu)} n^{1/2-m} (\log n)^{\frac{d-1}{d}} + \epsilon \| f \|_{L^2(D, \nu)} & \Lambda = \Lambda_n^{\text{HC}}
\end{cases}
\]

**Proof.** Since \( f \in L^2(D, \nu) \) we may let \( p = f_{\Lambda} \) be its orthogonal projection onto \( \text{span}\{\psi_n : n \in \Lambda\} \).

Then, using (A.1), (A.2) and (A.3), we obtain

\[
\| f - f_{\Lambda} \|_{L^\infty(D)} \leq \prod_{n \in \Lambda} \sum_{k=1}^{d} \sqrt{2n_k + 1} | \langle f, \psi_n \rangle_{L^2(D, \nu)} |
\]

\[
\leq \left( \sum_{n \notin \Lambda} \chi_{n,m}^\text{mix} | \langle f, \psi_n \rangle_{L^2(D, \nu)} |^2 \right)^{1/2} \left( \sum_{n \notin \Lambda} \frac{\prod_{k=1}^{d}(2n_k + 1)}{\chi_{n,m}^\text{mix}} \right)^{1/2},
\]

\[
\leq \| f \|_{H^m_{\text{mix}}(D, \nu)} \left( \sum_{n \notin \Lambda} \frac{\prod_{k=1}^{d}(2n_k + 1)}{\chi_{n,m}^\text{mix}} \right)^{1/2},
\]

where \( \chi_{n,m}^\text{mix} \) is as in (A.4). Observe that

\[
\chi_{n,m}^\text{mix} \geq c_{m,d} (\prod_{k=1}^{d} (n_k + 1))^{2m},
\]
for some constant $c_{m,d}$, and therefore

$$\|f - f_\Lambda\|_{L^\infty(D)} \leq c_{m,d} \|f\|_{H_{mix}^m(D,\nu)} \left( \sum_{n \notin \Lambda} \prod_{k=1}^d (n_k + 1)^{1-2m} \right)^{1/2}, \quad (5.5)$$

where here we also note that $\|f\|_{\tilde{H}_{mix}^m(D,\nu)} \leq c_{m,d} \|f\|_{H_{mix}^m(D,\nu)}$. We now specify the index set. First suppose that $\Lambda = \Lambda_{TP}^n$. Let $[d]$ denote the set of ordered tuples with entries in $\{1, \ldots, d\}$. Then

$$\sum_{n \notin \Lambda} \prod_{k=1}^d (n_k + 1)^{1-2m} = \sum_{\sigma \in [d]} \sum_{n_k=0}^n \sum_{k \notin \sigma} \prod_{k \in \sigma} (n_k + 1)^{1-2m}$$

$$= \sum_{\sigma \in [d]} \left( 1 + \sum_{l=0}^n \sum_{l \in \sigma} (n_l - 1)^{1-2m} \right) \left( \sum_{\sigma \in [d]} \sum_{l \notin \sigma} (n_l - 1)^{1-2m} \right) \leq c_{m,d} n^{2-2m}.$$ 

Substituting into (5.5) now gives the result for $\Lambda = \Lambda_{TP}^n$. Moreover, the result for the total degree index set now follows as well, after noting that $\Lambda_{TD}^n \supseteq \Lambda_{TP}^n / d$. Finally, for the hyperbolic cross index $\Lambda = \Lambda_{HC}^n$ set we use, for example, [1, Lem. 2.30] to get

$$\sum_{n \notin \Lambda} \prod_{k=1}^d (n_k + 1)^{1-2m} \leq c_{m,d} n^{2-2m} (\log(n))^{d-1},$$

as required.

It remains to consider the case where $\Omega$ is compactly contained in $D$. We first recall that univariate Legendre polynomials are uniformly bounded in compact subintervals of $(-1,1)$:

$$|\psi_n(y)| \leq c_r, \quad -1 + r \leq y \leq 1 - r, \quad \forall n \in \mathbb{N}_0,$$

for some $c_r > 0$. Hence $\|\phi_n\|_{L^\infty(\Omega)} \leq c_\Omega, \forall n \in \mathbb{N}_0^d$. Letting $p = f_\Lambda$ and arguing as above, we get

$$\|f - f_\Lambda\|_{L^\infty(\Omega)} \leq c_{m,d,\Omega} \|f\|_{H_{mix}^m(D,\nu)} \left( \sum_{n \notin \Lambda} \prod_{k=1}^d (n_k + 1)^{-2m} \right)^{1/2} \leq c_{m,d,\Omega} \|f\|_{H_{mix}^m(D,\nu)} \left( \sum_{n \notin \Lambda} \prod_{k=1}^d (n_k + 1)^{-2m} \right)^{1/2}$$

We now continue to argue as above, replacing the exponent $1 - 2m$ by $-2m$ throughout. \qed

6 Sample complexity

In this section we consider the efficiency, i.e. sample complexity, of the approximation. In view of Theorem 4.1 this can be analyzed by estimating the constant $C_{\Upsilon,\Lambda,\epsilon}$ defined in (4.2). Note that this depends on the choice of sample points $\Upsilon$ and index set $\Lambda$. Our main results are twofold. First, in §6.2 we show that when the sample points are drawn randomly and independently according to a suitable measure on $\Omega$ then the sample complexity can always be related to the constant of the Nikolskii-type inequality for the polynomial space $P_\Lambda$. Second, in §6.3 we show that for domains satisfying a suitable property this constant, and therefore the overall sample complexity, is at most quadratic in the dimension $N$ of the polynomial space $P_\Lambda$. 

18
6.1 The constant $C_{T,A}$

It is difficult to analyze $C_{T,A,\epsilon}$ directly, since it defined in terms of the singular values and singular vectors of the matrix $A$. In order to provide concrete bounds, we now consider

$$C_{T,A} = \sup \left\{ \|p\|_{L^2(\Omega,\mu)} : p \in P_A, \frac{1}{M} \sum_{y \in \mathcal{T}} |p(y)|^2 = 1 \right\}.$$ 

The connection between this constant and $C_{T,A,\epsilon}$ is as follows:

**Lemma 6.1.** Let $C_{T,A,\epsilon}$ be as in (4.2). Then $C_{T,A,\epsilon} \leq C_{T,A}$ and moreover $C_{T,A} = C_{T,A,\epsilon}$ whenever the minimum singular value of $A = A_{T,A}$ satisfies $\sigma_{\min}(A) > \epsilon$.

**Proof.** Since $C_{T,A,\epsilon}$ is the maximum of $C'_{T,A,\epsilon}$ and $C''_{T,A,\epsilon}$ we consider both of these constants separately. First consider $C'_{T,A,\epsilon}$. Let $b \in \mathbb{C}^M$, $\|b\|_2 = 1$, and notice that $T_A(A_{T,A,\epsilon})^*b \in P_A$. Hence, by the definition of $C_{T,A}$ and (4.5), we have

$$\left\| T_A(A_{T,A,\epsilon})^*b \right\|_{L^2(\Omega,\mu)} \leq C_{T,A} \left\| S_T T_A(A_{T,A,\epsilon})^*b \right\|_2.$$ 

Recall that $A = A_{T,A} = S_T T_A$. Therefore, by standard properties of the SVD, we have

$$\left\| S_T T_A(A_{T,A,\epsilon})^*b \right\|_2 \leq \|b\|_2 = 1,$$ 

and hence $\|T_A(A_{T,A,\epsilon})^*b\|_{L^2(\Omega,\mu)} \leq C_{T,A}$. Since $b$ was arbitrary we get $C'_{T,A,\epsilon} \leq C_{T,A}$.

On the other hand, suppose that $\sigma_{\min}(A) > \epsilon$. Let $p = T_A e \in P_A$. Since $A$ is full rank we have $b = S_T p \neq 0$. Without loss of generality, assume that $\|b\|_2 = 1$. Then

$$C''_{T,A,\epsilon} \geq \left\| T_A(A_{T,A,\epsilon})^* S_T A_{T,A} c \right\|_{L^2(\Omega,\mu)} = \left\| T_A(A_{T,A,\epsilon})^* A_{T,A} c \right\|_{L^2(\Omega,\mu)} = \|T_A c\|_{L^2(\Omega,\mu)} = \|p\|_{L^2(\Omega,\mu)},$$

where in the third step we use the fact that $A_{T,A}$ is full rank once more. Hence, since $p$ was arbitrary, we get $C''_{T,A,\epsilon} \geq C_{T,A}$, and therefore $C_{T,A,\epsilon} = C_{T,A}$ in this case.

Finally, consider $C''_{T,A,\epsilon}$. Let $d \in \mathbb{C}^N$, $\|d\|_2 = 1$. Then

$$\left\| T_A d - T_A(A_{T,A,\epsilon})^* T_A d \right\|_{L^2(\Omega,\mu)} \leq C_{T,A} \left\| S_T \left( T_A d - T_A(A_{T,A,\epsilon})^* A_{T,A} d \right) \right\|_2$$

$$= C_{T,A} \left\| \left( A_{T,A} - A_{T,\epsilon}(A_{T,A,\epsilon})^* A_{T,A} \right) d \right\|_2$$

$$= C_{T,A} \|U (\Sigma - \Sigma_{\epsilon}) V^* d\|_2 \leq C_{T,A} \epsilon \|d\|_2.$$ 

Since $d$ was arbitrary, we get $C''_{T,A,\epsilon} \leq C_{T,A}$ as required. On the other hand, if $\sigma_{\min}(A) > \epsilon$ then $\Sigma - \Sigma_{\epsilon} = 0$. Hence $C''_{T,A,\epsilon} = 0$. \qed

6.2 Random sampling for compact domains and Nikolskii-type inequalities

We now show that $C_{T,A}$ can be bounded using the constant of the Nikolskii-type inequality for the space $P_A \subset L^2(\Omega,\mu)$. To this end, let $N(P_A,\Omega,\mu)$ be the smallest positive number in the $(L^2(\Omega,\mu), L^\infty(\Omega))$ Nikolskii-type inequality

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

(6.1)

Then we have the following result:
**Theorem 6.2.** Let $0 < \delta, \gamma < 1$ and suppose that $y_1, \ldots, y_M$ are independent and randomly drawn according to the probability measure $\mu$ defined by (7.2). If

$$M \geq (N(P_\Lambda, \Omega, \mu))^2 \left( (1 - \delta) \log(1 - \delta) + \delta \right)^{-1} \log(N/\gamma),$$

where $N = |\Lambda|$ and $N(P_\Lambda, \Omega, \mu)$ is the constant of the Nikolskii-type inequality (6.1), then with probability at least $1 - \gamma$ the quantity $C_{T, \Lambda}$ satisfies

$$C_{T, \Lambda} \leq \frac{1}{\sqrt{1 - \delta}}.$$

**Proof.** Our proof is based on the essentially same arguments that have been used previously for tensor-product domains (see, for example, [21]). First, let $\{\Phi_n\}_{n \in \Lambda}$ be an orthonormal basis for $P_\Lambda$ in $L^2(\Omega, \mu)$. Let $p \in P_\Lambda$ be arbitrary and write $p = \sum_{n \in \Lambda} c_n \Phi_n$, so that

$$||p||_{L^2(\Omega, \mu)}^2 = \int_\Omega |p(y)|^2 d\mu(y) = ||c||_2^2,$$

where $c = (c_n)_{n \in \Lambda}$, and $\frac{1}{M} \sum_{y \in \mathcal{T}} |p(y)|^2 = c^* B c$, where $B \in \mathbb{C}^{N \times N}$ is the self-adjoint matrix with

$$(B)_{m,n} = \frac{1}{M} \sum_{y \in \mathcal{T}} \Phi_m(y) \Phi_n(y), \quad m, n \in \Lambda.$$

It follows that $C_{T, \Lambda} = 1/\sqrt{\lambda_{\min}(B)}$, where $\lambda_{\min}(B)$ is the minimal eigenvalue of $B$. We estimate this quantity by writing it in the usual way as the sum of independent random matrices. Write

$$B = \sum_{m=1}^M X_m, \quad X_m = \left\{ \frac{1}{M} \Phi_m(y) \Phi_n(y) \right\}_{m,n \in \Lambda}.$$

By construction, these matrices are independent, nonnegative definite and satisfy $\mathbb{E}(X_m) = \frac{1}{M} I$, where $I$ is the identity matrix. Moreover, for any $c \in \mathbb{C}^N$ we have

$$c^* X_m c = \frac{1}{M} \left| \sum_{n \in \Lambda} c_n \Phi_n(y_m) \right|^2 \leq \frac{(N(P_\Lambda, \Omega, \mu))^2}{M} \left\| \sum_{n \in \Lambda} c_n \Phi_n \right\|_{L^2(\Omega, \mu)}^2 = \frac{(N(P_\Lambda, \Omega, \mu))^2}{M} ||c||_2^2.$$

The Matrix Chernoff bound (see, for example, [43, Thm. 1.1]) now gives

$$\mathbb{P} \left( \lambda_{\min}(X) \leq (1 - \delta) \right) \leq N \exp \left( -\frac{(1 - \delta) \log(1 - \delta) + \delta}{M^{-1}(N(P_\Lambda, \Omega, \mu))^2} \right).$$

Setting the right-hand side equal to $\gamma$ and rearranging yields the result.

This leads to the following result on accuracy of the truncated SVD least-squares approximation:

**Corollary 6.3.** Let $0 < \delta, \gamma < 1$ and suppose that $y_1, \ldots, y_M$ are independent and randomly drawn according to the probability measure $\mu$ defined by (7.2). Let

$$M \geq (N(P_\Lambda, \Omega, \mu))^2 \left( (1 - \delta) \log(1 - \delta) + \delta \right)^{-1} \log(N/\gamma),$$

where $N = |\Lambda|$ and $N(P_\Lambda, \Omega, \mu)$ is the constant of the Nikolskii-type inequality (6.1). Then with probability at least $1 - \gamma$ the truncated SVD least-squares approximation $f_{T, \Lambda, \epsilon}$ of $f \in L^\infty(\Omega)$ satisfies

$$\|f - f_{T, \Lambda, \epsilon}\|_{L^2(\Omega, \mu)} \leq \left( 1 + \frac{1}{\sqrt{1 - \delta}} \right) E_{\Lambda, \epsilon}(f),$$

where $E_{\Lambda, \epsilon}(f)$ is as in (4.7).

**Remark 6.4** We note in passing that the above result is not specific to polynomials, but holds for any finite-dimensional subspace $P$ of $L^2(\Omega, \mu)$.
6.3 The λ-rectangle property and quadratic sample complexity

In view of Theorem 6.2, the sample complexity can be determined by estimating $N(P_\lambda, \Omega, \mu)$. Deriving bounds for this constant for general irregular domains in arbitrarily-many dimensions is a challenging open problem. We shall not attempt to resolve this problem in full generality here (see §9 for some further discussion). Instead, we now show that this constant can indeed be estimated for a large class of irregular domains whenever $\nu$ is the uniform measure.

The type of domain we now consider are those satisfying the so-called λ-rectangle property:

**Definition 6.5 (λ-rectangle property).** A compact domain $\Omega$ has the λ-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

$$\inf_{R \in \mathcal{R}} \text{Vol}(R) = \lambda \text{Vol}(\Omega).$$

There are many domains of interest that have this property. We now list several examples:

- **L-shaped domains.** These are unions of two rectangles, so clearly have this property.

- **Domains with linear constraints.** The domain

$$\Omega = \{-1 \leq y_1, y_2 \leq 1, y_1 + y_2 \leq 1\},$$

along with its various higher-dimensional generalizations, can be expressed as

$$\Omega = \bigcup_{x \in [0,1]} R_x, \quad R_x = [-1, x] \otimes [-1, 1 - x].$$

Hence it has the λ-rectangle property with $\lambda = 1/2$. Note that such domains can occur in problems such as surrogate forwards model construction in parameter studies; for instance, whenever two parameters $y_1$ and $y_2$, rather than being independent, satisfy a (possibly a priori unknown) linear relation.

- **Domains with exclusions.** The domain

$$\Omega = \{-1 \leq y_1, y_2 \leq 1, y_1^2 + y_2^2 \geq 1/2\},$$

along with various generalizations, satisfies the λ-rectangle property with $\lambda = \frac{9/4 - \sqrt{2}}{8 - \pi/2} \approx 0.13$. This follows after noting that

$$\Omega = \bigcup_{x \in [0,1/2]} (\pm [x, 1]) \otimes \left(\pm [\sqrt{1/4 - x^2}, 1]\right).$$

Note that such domains correspond to practical scenarios where, due to certain physical constraints, $f(y)$ can only be evaluated for $y$ not too close to zero.

See Figure 2 for illustrations. On the other hand, there are a number of notable domains that do not have this property. These include the unit Euclidean ball $\{y \in \mathbb{R}^d : \|y\|_2 \leq 1\}$ and the simplex $\{y \in \mathbb{R}^d : 0 \leq y_1, \ldots, y_{d-1} \leq 1, 0 \leq y_d \leq 1 - (y_1 + \ldots + y_{d-1})\}$. See §9 for additional details.
Theorem 6.6. Suppose that $\Omega \subseteq (-1,1)^d$ has the $\lambda$-rectangle property and let $P_\Lambda$ be constructed from the tensor Legendre polynomial basis on $(-1,1)^d$, where $\Lambda$ is any lower set (see Definition 3.1) of cardinality $|\Lambda| = N$. Let $\mu$ be the uniform probability measure on $\Omega$ and $N(P_\Lambda,\Omega,\mu)$ be the constant in the Nikolskii-type inequality (6.1). Then

$$ (N(P_\Lambda,\Omega,\mu))^2 \leq \frac{N^2}{\lambda}. $$

Proof. We first claim that $P_\Lambda = \text{span}\{\phi_n : n \in \Lambda\}$ coincides with the space

$$ P_\Lambda = \text{span}\{y \in \Omega \mapsto y^n : n = (n_1, \ldots, n_d) \in \Lambda\}. $$

Here we use the notation $y^n = y_1^{n_1} \cdots y_d^{n_d}$. Since $\phi_n(y)$ is a tensor Legendre polynomial we have

$$ \phi_n(y) = \prod_{k=1}^d \psi_{n_k}^{(k)}(y_k) = \prod_{k=1}^d \left( \sum_{m_k=0}^{n_k} a_{m_k,n_k} y_k^{n_k} \right) = \sum_{m_1=0}^{n_1} \cdots \sum_{m_d=0}^{n_d} a_{m,n} y^m \in P_\Lambda, $$

where $a_{m_k,n_k}$ are the coefficients of $\psi_{n_k}^{(k)}$ in the monomial basis and $a_{m,n} = \prod_{k=1}^d a_{m_k,n_k}$. Since $m_k \leq n_k$ for all $k$, it follows from the lower set assumption that $m \in \Lambda$ and therefore $\phi_n \in P_\Lambda$. Hence $P_\Lambda \subseteq P_\Lambda$. In an identical manner, one also finds that $y \mapsto y^n$ is in $P_\Lambda$, and therefore $P_\Lambda \subseteq P_\Lambda$, as required.

Now let $p \in P_\Lambda$ and $y \in \Omega$ with $y \in R$ for some $R \in \mathcal{R}$. Define the uniform measure on $R$ as

$$ d\tilde{\mu}(y) = \frac{1}{\text{Vol}(R)} \, dy, $$

and note that $|p(y)| \leq N(P_\Lambda, R, \tilde{\mu}) \|p\|_{L^2(R,\tilde{\mu})}$, where $N(P_\Lambda, R, \tilde{\mu})$ is the Nikolskii constant for the space $P_\Lambda$ with respect to $L^2(R,\tilde{\mu})$. It is known that $(N(P_\Lambda, R, \tilde{\mu}))^2 \leq N^2$ [32] Thm. 6]. Also

$$ \|p\|^2_{L^2(R,\tilde{\mu})} = \frac{1}{\text{Vol}(R)} \int_R |p(y)|^2 \, dy \leq \frac{\text{Vol}(\Omega)}{\text{Vol}(R)} \int_\Omega |p(y)|^2 \, d\tilde{\mu}(y) \leq \frac{1}{\lambda} \|p\|^2_{L^2(\Omega,\mu)}. $$

Hence $|p(y)|^2 \leq \frac{N^2}{\lambda} \|p\|^2_{L^2(\Omega,\mu)}$. Since $y \in \Omega$ and $p \in P_\Lambda$ were arbitrary, we now get the result. \hfill $\square$

Combining this with Corollary 5.3 now immediately gives the following result:

Corollary 6.7. Suppose that $\Omega \subseteq (-1,1)^d$ has the $\lambda$-rectangle property and let $P_\Lambda$ be constructed from the tensor Legendre polynomial basis on $(-1,1)^d$, where $\Lambda$ is any lower set of cardinality...
7 Truncated estimators and $L^2$-error bounds

The error bounds proved in Corollary 6.3 and elsewhere have the limitation of relating (in probability) the $L^2$-norm of the error to an approximation error $E_{\lambda,\epsilon}(f)$ measured in the $L^\infty$-norm. In this penultimate section we show that it is possible to bound the $L^2$-norm of a related estimator in expectation using the $L^2$-norm approximation error

$$E_{\lambda,\epsilon}(f) = \inf \left\{ \| f - p \|_{L^2(\Omega,\mu)} + \epsilon \| p \|_{L^2(D,\nu)} : p \in P_\lambda \right\}. \quad (7.1)$$

We follow the approach of [21]. First, suppose that $f \in L^\infty(\Omega,\mu)$ and let $L \geq 0$ be such that $\| f \|_{L^\infty(\Omega)} \leq L$. Now define the truncation operator

$$T_L(g)(y) = \text{sign}(g(y)) \min\{ |g(y)|, L \},$$

where sign($z$) denotes the complex sign of $z \in \mathbb{C}$. If $f_{\lambda,\epsilon}$ is the truncated SVD least-squares approximation we now define the new approximation

$$f_{\lambda,\epsilon,L} = T_L(f_{\lambda,\epsilon}). \quad (7.2)$$

Our main result is now the following:

**Theorem 7.1.** Let $0 < \delta, \gamma < 1$ and $f \in L^\infty(\Omega)$ with $\| f \|_{L^\infty(\Omega)} \leq L$ for some $L \geq 0$. Let $y_1, \ldots, y_M$ be independent and randomly drawn according to $\mu$ and $f_{\lambda,\epsilon,L}$ be as in (7.2). If $E_{\lambda,\epsilon}(f)$ is as in (7.1) and

$$M \geq (N(P_\lambda,\Omega,\mu))^2((1 - \delta) \log(1 - \delta) + \delta)^{-1} \log(N/\gamma), \quad (7.3)$$

where $N = |\Lambda|$ and $N(P_\lambda,\Omega,\mu)$ is the constant of the Nikolskii-type inequality (6.7), then

$$\mathbb{E} \left( \| f - f_{\lambda,\epsilon,L} \|_{L^2(\Omega,\mu)}^2 \right) \leq \frac{2 - \delta}{1 - \delta} \left( \mathbb{E}_{\lambda,\epsilon}(f) \right)^2 + 4L^2 \gamma. \quad (7.4)$$

**Proof.** The proof is based on [21] Thm. 2. Let $E$ be the event $C_{\lambda,\epsilon} \leq \frac{1}{\sqrt{1 - \delta}}$, where $C_{\lambda,\epsilon}$ is as in (4.2). Lemma 6.1, Theorem 6.2 and the measurement condition (7.3) give that $\mathbb{P}(E^c) \leq \gamma$. Now let $d\mu$ be the uniform measure on $\Omega$ and $d\mu_M = d\mu \otimes \cdots \otimes d\mu$ be the probability measure of the draw $y_1, \ldots, y_M$. Then

$$\mathbb{E} \left( \| f - f_{\lambda,\epsilon,L} \|_{L^2(\Omega,\mu)}^2 \right) d\mu_M = \int_E \| f - f_{\lambda,\epsilon,L} \|_{L^2(\Omega,\mu)}^2 d\mu_M + \int_{E^c} \| f - f_{\lambda,\epsilon,L} \|_{L^2(\Omega,\mu)}^2 d\mu_M \leq \int_E \| f - f_{\lambda,\epsilon,L} \|_{L^2(\Omega,\mu)}^2 d\mu_M + 4L^2 \gamma. \quad (7.4)$$
It remains to bound the first term. Assume the event $E$ occurs and let $p \in P_\Lambda$ be such that
\[ \|f - p\|^2_{L^2(\Omega, \mu)} + \epsilon \|p\|^2_{L^2(\Omega, \mu)} = \tilde{E}_{\Lambda, \epsilon}(f) \] (it is straightforward to show that such a minimizer exists, since $P_\Lambda$ is finite dimensional). Then, arguing as in the proof of Theorem 4.1 and using the fact that $C_{\Upsilon, \Lambda, \epsilon} \leq \frac{1}{\sqrt{1 - \delta}}$, we have
\[
\|f - f_{\Upsilon, \Lambda, \epsilon}\|_{L^2(\Omega, \mu)}^2 \leq \left(\|f - p\|_{L^2(\Omega, \mu)} + C_{\Upsilon, \Lambda, \epsilon}^2 \|\Upsilon(f - p)\|_2 + \epsilon C_{\Upsilon, \Lambda, \epsilon}^2 \|p\|_{L^2(\Omega, \mu)}\right)^2
\leq 3 \|f - p\|_{L^2(\Omega, \mu)}^2 + \frac{3}{1 - \delta} \|\Upsilon(f - p)\|_2^2 + \frac{3\epsilon^2}{1 - \delta} \|p\|_{L^2(\Omega, \mu)}^2.
\]

Hence
\[
\int_E \|f - f_{\Upsilon, \Lambda, \epsilon}\|_{L^2(\Omega, \mu)}^2 d\mu \leq 3 \int_E \|f - p\|_{L^2(\Omega, \mu)}^2 d\mu + \frac{3}{1 - \delta} \mathbb{E}\left(\|\Upsilon(f - p)\|_2^2\right) + \frac{3\epsilon^2}{1 - \delta} \mathbb{E}\|p\|_{L^2(\Omega, \mu)}^2.
\]
Observe that $\mathbb{E}\left(\|\Upsilon(f - p)\|_2^2\right) = \mathbb{E}|f(y) - p(y)|^2 = \|f - p\|^2_{L^2(\Omega, \mu)}$. Therefore we obtain
\[
\int_E \|f - f_{\Upsilon, \Lambda, \epsilon}\|_{L^2(\Omega, \mu)}^2 d\mu \leq 3 \frac{2 - \delta}{1 - \delta} \left(\|f - p\|_{L^2(\Omega, \mu)}^2 + \epsilon^2 \|p\|_{L^2(\Omega, \mu)}^2\right) \leq 3 \frac{2 - \delta}{1 - \delta} \left(\tilde{E}_{\Lambda, \epsilon}(f)\right)^2.
\]
Substituting this into (7.4) now gives the result. \hfill \Box

Much like in §3, we can establish bounds for $\tilde{E}_{\Lambda, \epsilon}(f)$ under different regularity conditions:

**Theorem 7.2.** Let $P_\Lambda$ be constructed from the tensor Legendre polynomial basis on $L^2(D, \nu)$, where $D = (-1,1)^d$, $\nu$ is the uniform measure on $D$, and $\Lambda = \Lambda_n$ is either the tensor product (3.4) or total degree (3.5) index set of degree $n$. If $\Omega \subseteq D$ and $f \in H^m(D, \nu)$ for some $m \geq 1$, then
\[
\tilde{E}_{\Lambda, \epsilon}(f) \leq c_{m, d} \|f\|_{H^m(D, \nu)} n^{-m} + \|f\|_{L^2(D, \nu)}.
\]
Conversely, if $\Omega \subseteq (-1,1)^d$ is Lipschitz and $f \in H^m(\Omega, \mu)$ where $\mu$ is the uniform measure on $\Omega$ and $m \geq 1$, then
\[
\tilde{E}_{\Lambda, \epsilon}(f) \leq c_{m, d, \Omega} \left(n^{-m} + \epsilon\right) \|f\|_{H^m(\Omega, \mu)}.
\]
Finally, if $\Lambda = \Lambda_n^{HC}$ is the hyperbolic cross index set (3.4) then the same results hold with $n^{-m}$ replaced by $n^{-m/d}$.

**Proof.** As in the proof of Theorem 5.1, if $f \in H^m(D, \nu)$ we let $f_\Lambda$ be the orthogonal projection of $f$ onto $\text{span}\{\psi_n : n \in \Lambda\} \subseteq L^2(D, \nu)$. Then by Parseval’s identity and (A.3),
\[
\|f - f_\Lambda\|^2_{L^2(D, \nu)} = \sum_{n \notin \Lambda} |\langle f, \psi_n \rangle_{L^2(D, \nu)}|^2 \leq \frac{1}{\min_{n \notin \Lambda} \{\chi_{n, m}\}} \sum_{n \notin \Lambda} \chi_{n, m} |\langle f, \psi_n \rangle_{L^2(D, \nu)}|^2 \\
\leq \frac{1}{\min_{n \notin \Lambda} \{\chi_{n, m}\}} \|f\|^2_{H^m(D, \nu)}.
\]
It remains to bound $\min_{n \notin \Lambda} \{\chi_{n, m}\}$ for the three index sets. Using (A.4), we first observe that
\[
\chi_{n, m} = \sum_{|\beta| \leq m} \prod_{k=1}^d (n_k (n_k + 1))^{j_k} \geq |n|^2_{\infty} > n^{2m}, \quad n \notin \Lambda_{\text{TP}}^{\text{TP}}.
\]
Similarly, for the total degree index set
\[
\chi_{n, m} \geq c_{m, d} |n|^2_{1} > c_{m, d} n^{2m}, \quad n \notin \Lambda_{\text{TD}}^{\text{TD}},
\]
and
and for the hyperbolic cross
\[ \chi_{n,m} \geq c_{m,d} |n|_{\text{hc}}^{2m/d} > c_{m,d} n^{2m/d}, \quad n \notin \Lambda_n^{\text{HC}}. \]

This gives the first result. For the second result, we argue as in the proof of Theorem 5.1 to construct an extension \( g \in H^m(D, \nu) \) of \( f \), and then use the previously-derived bounds.

**Theorem 7.3.** Let \( P_\Lambda \) be constructed from the tensor Legendre polynomial basis on \( L^2(D, \nu) \), where \( D = (-1,1)^d \) and \( \nu \) is the uniform measure on \( D \). If \( f \in H^m_{\text{mix}}(D, \nu) \) for some \( m \geq 1 \) then
\[ \tilde{E}_{\Lambda, \epsilon}(f) \leq c_{m,d} \|f\|_{H^m_{\text{mix}}(D, \nu)} n^{-m} + \epsilon \|f\|_{L^2(D, \nu)}, \]

when \( \Lambda = \Lambda_n^{\text{TP}}, \Lambda = \Lambda_n^{\text{TD}} \) or \( \Lambda = \Lambda_n^{\text{HC}} \).

**Proof.** Consider the setup of the previous proof. We have
\[ \|f - f_\Lambda\|^2_{L^2(D, \nu)} \leq \frac{1}{\min_{n \notin \Lambda} \chi_{n,m}^{\text{mix}}} \|f\|^2_{H^m_{\text{mix}}(D, \nu)}, \]

where \( \chi_{n,m}^{\text{mix}} \) is as in \( \text{(A.4)} \). We now observe that
\[ \chi_{n,m}^{\text{mix}} = \sum_{|j|_{\infty} \leq m} \prod_{k=1}^d (n_k(n_k + 1))^{j_k} \geq c_{m,d} |n|_{\text{hc}}^{2m} > c_{m,d} n^{2m}, \quad n \notin \Lambda, \]

where \( \Lambda \) is any of the three index sets consider. The result now follows immediately. \( \square \)

## 8 Numerical results

We conclude this paper with several numerical experiments illustrating the theoretical results. Unless otherwise stated we use Legendre polynomials on \( D = (-1,1)^d \), hyperbolic cross index sets, samples drawn independently from the uniform measure on \( \Omega \) and a threshold parameter \( \epsilon = 10^{-8} \).

### 8.1 Function regularity

First, we consider approximation of several bivariate functions. The left panel of Fig. 3 shows the approximation of a smooth function on an annulus. Note that the function is singular on \( D \setminus \Omega \). However, as predicted by the results of §5 this does not hamper the approximation on \( \Omega \). The right panel shows the approximation of a function defined on the Mandelbrot set. This domain is not Lipschitz, but since the function has a smooth extension to the whole of \( D \), an accurate approximation is obtained. This also agrees with the results of §5. Note that in neither case does the domain need to be known in advance in order to compute the approximation. It is defined implicitly by the data.

### 8.2 Sample complexity

In Figure 4 we examine the sample complexity of polynomial frame approximations for a two-dimensional circular domain. This involves computing the constant \( C_{T, \Lambda, \epsilon} \), which is discussed in the remark below. Fig. 4(a) suggests that quadratic oversampling is sufficient in this case, even though the domain is not of \( \lambda \)-rectangle type. Moreover, linear or log-linear oversampling results to exponential increase in \( C_{T, \Lambda, \epsilon} \), up to roughly \( 1/\epsilon \) (recall that \( C_{T, \Lambda, \epsilon} \leq 1/\epsilon \); see §4.3). On the other
$$f(y_1, y_2) = \log(8(y_1^2 + y_2^2)) - 2(y_1^2 + y_2^2)$$
$$\Omega = \{-1 \leq y_1, y_2 \leq 1 : f(y_1, y_2) \geq 0\}$$
$$\Lambda = \Lambda^{HC}_{200}$$

Figure 3: Pointwise error for polynomial frame approximations over two bivariate domains.

hand, Figs. 4(b),(c) suggest that log-linear oversampling is sufficient whenever domain $\Omega$ does not touch the bounding cube $D$. This interesting phenomenon, which is at odds with the quadratic rates predicted in §6, has been shown in the one dimensional-setting through extensive numerical experiments on the Fourier basis with deterministic equally-spaced points \[7, 9\].

While we currently have no proof of this observation, it is possible to give an intuitive explanation. The sample complexity is dictated by the maximal growth of a polynomial (in an $L^2$-sense) on $\Omega$, when it is bounded at $M$ points in $\Omega$. Polynomials that grow large in this sense must also be large outside $\Omega$. That is, it has large coefficients when represented in the Legendre basis. Yet, when regularizing via the truncated SVD, such polynomials are excluded from the resulting approximation space, thus in effect lowering the required sample complexity. This also explains why the sample complexity appears to change when $r$ is varied: for $r = 1$ the boundaries of $\Omega$ and $D$ intersect, but this ceases to be the case once $r < 1$. Formalizing this intuition into a proof is an open problem.

Remark 8.1 As shown in \[6\], the constants $C'_{T,\Lambda,\epsilon}$ and $C''_{T,\Lambda,\epsilon}$ can be expressed as
\begin{equation}
C'_{T,\Lambda,\epsilon} = \sqrt{\lambda_{\max}((B')^*GB')}, \quad C''_{T,\Lambda,\epsilon} = \epsilon^{-1} \sqrt{\lambda_{\max}((B'')^*GB'')},
\end{equation}
where $G = G^\Lambda$ is the Gram matrix of the truncated frame \[2.3\], $B' = (A^T,\Lambda,\epsilon)^\dagger = V (\Sigma^\epsilon)^\dagger U^*$ and $B'' = V I_{\epsilon}^\perp V^*$. Here $U \Sigma V^*$ is the SVD of $A$, and $I_{\epsilon}^\perp$ is the diagonal matrix with the $n$th entry 1 if $\sigma_n \leq \epsilon$ and zero otherwise. Computing the Gram matrix $G$ over irregular domains is difficult, but it can be approximated by Monte–Carlo integration. That is, $G \approx H^*H$ where
\begin{equation}
H = H_{K,\Lambda} = \left( \frac{1}{\sqrt{K}} \phi_n(z_k) \right)_{k=1,\ldots,K,n\in\Lambda} \in \mathbb{C}^{K\times N},
\end{equation}
and $z_1, \ldots, z_K$ are drawn independently from the uniform measure on $\Omega$. Replacing $G$ by $H^*H$ in \[8.1\] and using standard properties of singular values leads to the simpler approximate expressions
\begin{equation}
C'_{T,\Lambda,\epsilon} \approx \left\|HV (\Sigma^\epsilon)^\dagger\right\|_2, \quad C''_{T,\Lambda,\epsilon} \approx \epsilon^{-1} \left\|HV I_{\epsilon}^\perp\right\|_2
\end{equation}
where $\|\cdot\|_2$ denotes the matrix 2-norm.
Figure 4: The constant $C_{T, A, \epsilon}$ against $N$ for various different choices of $M$. The domain $\Omega$ is a circle of radius $r$ in $d = 2$ dimensions. Computations were averaged over 20 trials with the median value taken. The constant $C_{T, A, \epsilon}$ was computed as in Remark 8.1 using a precomputed grid of $K = 10000$ Monte–Carlo points in $\Omega$.

Figure 5: The median error over 20 trials versus $M$ for approximating the function $f(\mathbf{y}) = 1/\sum_{i=1}^{d} \sqrt{|y_i|}$ on the annular domain $\Omega = \{ \mathbf{y} : r/4 \leq \|\mathbf{y}\|_2 \leq r \}$. For each $M$, the value of $N$ is chosen as the largest such that $N \log(N) \leq M$.

### 8.3 Higher dimensions

In Fig. 5, we consider approximation error for various different dimensions. This figure shows the approximation error versus $M$ for an annular region of several different radii. Log-linear oversampling was used throughout. It is noticeable that when $r = 1$, meaning that $\Omega$ touches the boundary of $D$, the approximation is ill-conditioned, and the error duly increases for large enough $M$. As is to be expected, this increase is most severe in lower dimensions (since the cardinality of the polynomial space is largest in this setting). Conversely, as soon as $\Omega$ is compactly contained in $D$, the approximation error decreases as $M$ increases. Note that the function being approximated is smooth in $\Omega$ but singular at $\mathbf{y} = \mathbf{0} \in D \setminus \Omega$. As predicted by the results of §5, the approximation error decreases rapidly despite this singularity.

### 8.4 Other bases

Finally, in Figs. 6 & 7, we consider the use of different orthogonal bases on the extended domain. First, in Fig. 7, we consider the tensor cosine basis defined on $D = (-T, T)^d$, where $T \geq 1$ is a parameter. The basis elements in this case are tensor-products on the univariate functions $$\phi_n(y) = \cos(n\pi(y + T)/(2T)), \quad n = 0, 1, 2, \ldots.$$
Figure 6: The median error over 20 trials versus $M$ for approximating the function $f(y) = \exp\left(\sum_{i=1}^{d} y_i/d\right)$ on the corner domain $\Omega = \{y \in (-1,1)^d : y_1 + \ldots + y_d \leq 1\}$ using the tensor cosine basis on $[-T,T]^d$. For each $M$, the value of $N$ is chosen as the largest such that $N \log(N) \leq M$.

Figure 7: The median error over 20 trials versus $M$ for approximating the Genz product peak function $f(y) = \prod_{i=1}^{d} \frac{y_i}{d+\frac{d}{4}+(y_i+(-1)^{i+1})(i+1)^2}$ on the domains $\Omega = \{y \in (-1,1)^d : y_1 + \ldots + y_d \leq 1\}$ (left), $\Omega = \{y \in (-1,1)^d : \|y\|_2 \geq 1/2\}$ (middle) and $\Omega = \{y \in (-1,1)^d : \|y\|_2 \leq 1\}$ (right) using Chebyshev polynomials on $D = (-1,1)^d$ and samples drawn randomly according to the Chebyshev measure restricted to $\Omega$. For each $M$, the value of $N$ is chosen as the largest such that $N \log(N) \leq M$.

When $T = 1$, the domain $\Omega$ is not compactly contained in $D$ and the approximation error decreases slowly, at rate of $N^{-1}$. This stems from the fact that cosine expansions, much like Fourier expansions, only converge rapidly for smooth functions that satisfy additional boundary conditions [1]. When there is no gap between $\Omega$ and the boundary of $D$, there are no smooth extensions of $f$ satisfying these boundary conditions. Conversely, once $T > 1$ and $\Omega$ is compactly-contained in $D$, such extensions exists, and we witness correspondingly faster convergence. Error estimates similar to those proved in [5] can also be established for these approximations. See [2] for further details.

In Fig. 7 we consider Chebyshev polynomials on $D = (-1,1)^d$ and random sampling according to the tensor Chebyshev density restricted to $\Omega$. Unlike for Legendre polynomials (see Fig. 5) the approximation converges for all domains $\Omega$. This is perhaps surprising. Points drawn on a cube according to the Chebyshev density cluster quadratically near the boundary of the cube, thus permitting a lower sample complexity. However, points drawn according to the same density when restricted to a subdomain $\Omega$ do not necessarily cluster in this way over the whole boundary of $\Omega$. Yet, in spite of this we still witness rapid convergence.
9 Conclusions and challenges

In this work, we have provided a theoretical analysis of polynomial approximations of multivariate functions on irregular domains. In particular, for functions of mixed Sobolev regularity, we have shown that the regularized least-squares polynomial frame approximation is well-conditioned and converges algebraically down to a given threshold parameter $\epsilon$. Moreover, for a large class of domains, the sample complexity is provably quadratic in the dimension of the approximation space.

This paper marks only a first foray into the topic of multivariate polynomial approximation on irregular domains. Consequently, there are a number of significant directions for future research. We now highlight three such areas:

1. **Sample complexity estimates.** When sampling from the uniform measure, this paper has shown quadratic sample complexity for $\lambda$-rectangle domains. This rate is optimal. However, as mentioned, many domains do not have this property. We conjecture that the same sample complexity holds for a much more general class of domains which includes spheres and simplices (two notable domains which do not have the $\lambda$-rectangle property) and which is potentially also invariant under rotations (rotations generally destroy the $\lambda$-rectangle property). However, this remains an open problem. As discussed in §8, linear sample complexity is sufficient whenever $\Omega$ is compactly contained in $D$. While there is intuition behind this observation, we currently have no proof.

2. **Optimal sampling.** In tensor-product domains, recent work has identified densities for random sampling which achieve near-optimal log-linear sample complexities for least-squares approximations [24]. Extending these densities to irregular domains is an open problem. A singular challenge is the need to derive densities from which samples can be easily drawn. In [24] the tensor-product structure is heavily leveraged for this purpose; an option which is not available in irregular domains. More generally, determining efficient sampling densities for irregular domains is an open problem.

3. **Compressed sensing-based polynomial approximations.** Polynomial-based compressed sensing approaches have recently proved effective for high-dimensional approximation in regular domains (see [3, 4, 19, 35, 37, 46] and references therein). A problem for future work is to extend these approaches to irregular domains. Note that since polynomial frames are redundant, the usual compressed sensing theory for orthogonal bases does not apply.

Acknowledgements

A preliminary version of this work was presented at the Oberwolfach conference on “Multiscale and High-Dimensional Problems”. The authors would like to thank the organizers and participants for the useful discussions and feedback received during the conference. They would also like to thank Claudio Canuto, Albert Cohen, Paul Constantine, Ingrid Daubechies, Bert Deburner, Alireza Doostan, Wolfgang Hackbusch, Vladimir Temlyakov and Tino Ullrich. The first author is supported by NSERC grant 611675, as well as an Alfred P. Sloan Research Fellowship. The second author is supported by FWO-Flanders projects G.0641.11 and G.A004.14, as well as by KU Leuven project C14/15/055.

A Background on Legendre polynomials

This section contains some ancillary results on Legendre polynomials used earlier in the paper. Let $\{\psi_n\}_{n=0}^\infty$ be the orthonormal Legendre polynomial basis with respect to the uniform measure on
\((-1, 1)\). This is defined by
\[ \psi_n(y) = \sqrt{2n + 1} P_n(y), \]  
(A.1)
where \(P_n\) is the classical Legendre polynomial with normalization \(P_n(1) = 1\).

### A.1 One dimensional Legendre–Sobolev spaces

Recall that \(\psi_n(y)\) are the eigenfunctions of the Sturm–Liouville operator \(\mathcal{L}\), defined by
\[ \mathcal{L}f(y) = ((1 - y^2)f'(y))'. \]
Specifically, \(\mathcal{L}\psi_n(y) = n(n + 1)\psi_n(y)\). The operator \(\mathcal{L}\) is compact, self-adjoint and nonnegative definite. Indeed, we note that
\[ \langle \mathcal{L} f, g \rangle_{L^2(D, \nu)} = \langle f', g' \rangle_{L^2(D, \rho)} = \langle f, \mathcal{L} g \rangle_{L^2(D, \nu)}. \]
where \(D = (-1, 1)\), \(\nu\) is the uniform measure on \((-1, 1)\) and \(d\rho(y) = \frac{1 - y^2}{2} dy\). It therefore has a well-defined square root \(S = L^{1/2}\). This operator satisfies
\[ \|S f\|_{L^2(D, \nu)}^2 = \|f'\|_{L^2(D, \rho)}^2 = \langle \mathcal{L} f, f \rangle_{L^2(D, \nu)}. \]

With this in hand, for \(j \in \mathbb{N}\) let \(S^j = S \circ S \circ \cdots \circ S\) be the \(j\)-fold composition of \(S\) and define the Legendre–Sobolev space
\[ H^m(D, \nu) = \{ f \in L^2(D, \nu) : S^j f \in L^2(D, \nu), \ j = 0, \ldots, m \}, \]
with inner product and norm
\[ \langle f, g \rangle_{H^m(D, \nu)} = \sum_{j=0}^{m} \langle S^j f, S^j g \rangle_{L^2(D, \nu)}, \quad \|f\|_{H^m(D, \nu)} = \sqrt{\sum_{j=0}^{m} \|S^j f\|_{L^2(D, \nu)}^2}. \]

The set \{\(\psi_n\)\}_{n \in \mathbb{N}_0} is an orthogonal basis for \(H^m(D, \nu)\), and one has the expression
\[ \|f\|_{H^m(D, \nu)} = \sqrt{\sum_{n=0}^{\infty} \chi_{n,m} \|\langle f, \psi_n \rangle_{L^2(D, \nu)}\|^2}, \quad \chi_{n,m} = \sum_{j=0}^{m} (n(n + 1))^j. \]

### A.2 Multidimensional Legendre–Sobolev spaces

Let \(D = (-1, 1)^d\) be the unit cube and define the tensor Legendre polynomial basis \{\(\psi_n\)\}_{\mathbb{N}^d_0} as
\[ \psi_n(y) = \prod_{k=1}^{d} \psi_{n_k}(y_k), \quad n = (n_1, \ldots, n_d) \in \mathbb{N}^d_0, \ y = (y_1, \ldots, y_d) \in D. \]  
(A.2)

For \(k = 1, \ldots, d\), let \(\mathcal{L}_k\) be the compact, self-adjoint nonnegative definite operator
\[ \mathcal{L}_k f(y) = \frac{\partial}{\partial y_k} \left( (1 - y_k^2) \frac{\partial f}{\partial y_k} \right), \]
with corresponding square-root \(S_k = \mathcal{L}_k^{1/2}\) and powers \(S^j_k = S_k \circ \cdots \circ S_k\). Now let \(j = (j_1, \ldots, j_d) \in \mathbb{N}^d_0\) be a multi-index. We define the operator
\[ S^j = S_{j_1}^{j_1} \circ \cdots \circ S_{j_d}^{j_d}. \]
With this in hand, we now introduce the $d$-dimensional Legendre–Sobolev spaces

$$\tilde{H}^m(D, \nu) = \left\{ f \in L^2(D, \nu) : S^j f \in L^2(D, \nu), \ |j|_1 \leq m \right\},$$

with inner product and norm

$$\langle f, g \rangle_{\tilde{H}^m(D, \nu)} = \sum_{|j|_1 \leq m} \langle S^j f, S^j g \rangle_{L^2(D, \nu)}, \quad \|f\|_{\tilde{H}^m(D, \nu)} = \sqrt{\sum_{|j|_1 \leq m} \|S^j f\|_{L^2(D, \nu)}^2}.$$ 

We also introduce the mixed $d$-dimensional Legendre–Sobolev spaces as

$$\tilde{H}^m_{\text{mix}}(D, \nu) = \left\{ f \in L^2(D, \nu) : S^j f \in L^2(D, \nu), \ |j|_\infty \leq m \right\},$$

with inner product and norm

$$\langle f, g \rangle_{\tilde{H}^m_{\text{mix}}(D, \nu)} = \sum_{|j|_\infty \leq m} \langle S^j f, S^j g \rangle_{L^2(D, \nu)}, \quad \|f\|_{\tilde{H}^m_{\text{mix}}(D, \nu)} = \sqrt{\sum_{|j|_\infty \leq m} \|S^j f\|_{L^2(D, \nu)}^2}.$$ 

Both these norms can be characterized in terms of Legendre polynomial coefficients. Specifically,

$$\|f\|_{\tilde{H}^m(D, \nu)} = \sqrt{\sum_{n \in \mathbb{N}_0^d} \chi_{n,m} \langle f, \psi_n \rangle_{L^2(D, \nu)}^2}, \quad \|f\|_{\tilde{H}^m_{\text{mix}}(D, \nu)} = \sqrt{\sum_{n \in \mathbb{N}_0^d} \chi_{n,m}^{\text{mix}} \langle f, \psi_n \rangle_{L^2(D, \nu)}^2},$$

where

$$\chi_{n,m} = \sum_{|j|_1 \leq m} \prod_{k=1}^d \left( n_k(n_k + 1) \right)^{j_k}, \quad \chi_{n,m}^{\text{mix}} = \sum_{|j|_\infty \leq m} \prod_{k=1}^d \left( n_k(n_k + 1) \right)^{j_k}.$$ 

(A.3)

Finally, we note in passing that one has the continuous embeddings $H^m(D, \nu) \hookrightarrow \tilde{H}^m(D, \nu)$ and $H^m_{\text{mix}}(D, \nu) \hookrightarrow \tilde{H}^m_{\text{mix}}(D, \nu)$.

References

[1] B. Adcock. Modified Fourier expansions: theory, construction and applications. PhD thesis, University of Cambridge, 2010.

[2] B. Adcock. Multivariate modified Fourier series and application to boundary value problems. Numer. Math., 115(4):511–552, 2010.

[3] B. Adcock. Infinite-dimensional compressed sensing and function interpolation. Found. Comput. Math. (to appear), 2017.

[4] B. Adcock, S. Brugiapaglia, and C. G. Webster. Polynomial approximation of high-dimensional functions via compressed sensing. arXiv:1703.06987, 2017.

[5] B. Adcock and D. Huybrechs. Frames and numerical approximation. arXiv:1612.04464, 2016.

[6] B. Adcock and D. Huybrechs. Frames and stable numerical approximation from indirect data. In preparation, 2016.

[7] B. Adcock, D. Huybrechs, and J. Martín-Vaquero. On the numerical stability of Fourier extensions. Found. Comput. Math., 14(4):635–687, 2014.

[8] B. Adcock, R. Platte, and A. Shadrin. Optimal sampling rates for approximating analytic functions from pointwise samples. arXiv:1610.04769, 2016.
[9] B. Adcock and J. Ruan. Parameter selection and numerical approximation properties of Fourier extensions from fixed data. *J. Comput. Phys.*, 273:453–471, 2014.

[10] N. Albin and O. P. Bruno. A spectral FC solver for the compressible Navier–Stokes equations in general domains I: Explicit time-stepping. *J. Comput. Phys.*, 230(16):6248–6270, 2011.

[11] J. Boyd. Fourier embedded domain methods: extending a function defined on an irregular region to a rectangle so that the extension is spatially periodic and $C^\infty$. *Appl. Math. Comput.*, 161(2):591–597, 2005.

[12] J. P. Boyd. A comparison of numerical algorithms for Fourier Extension of the first, second, and third kinds. *J. Comput. Phys.*, 178:118–160, 2002.

[13] O. Bruno and M. Lyon. High-order unconditionally stable FC-AD solvers for general smooth domains I. Basic elements. *J. Comput. Phys.*, 229(6):2009–2033, 2010.

[14] O. P. Bruno, Y. Han, and M. M. Pohlman. Accurate, high-order representation of complex three-dimensional surfaces via Fourier continuation analysis. *J. Comput. Phys.*, 227(2):1094–1125, 2007.

[15] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numer.*, 13:147–269, 2004.

[16] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang. *Spectral methods: Fundamentals in Single Domains*. Springer, 2006.

[17] A. Chkifa, A. Cohen, G. Miglierati, F. Nobile, and R. Tempone. Discrete least squares polynomial approximation with random evaluations - application to parametric and stochastic elliptic pdes. *ESAIM Math. Model. Numer. Anal.*, 49(3):815–837, 2015.

[18] A. Chkifa, A. Cohen, and C. Schwab. Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs. *J. Math. Pures Appl.*, 103:400–428, 2015.

[19] 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.

[20] O. Christensen. *An Introduction to Frames and Riesz Bases*. Birkhauser, 2003.

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

[22] A. Cohen and R. A. DeVore. Approximation of high-dimensional parametric PDEs. *Acta Numer.*, 24:1–159, 2015.

[23] A. Cohen, R. A. DeVore, and C. Schwab. Convergence rates of best $N$-term Galerkin approximations for a class of elliptic sPDEs. *Found. Comput. Math.*, 10:615–646, 2010.

[24] A. Cohen and G. Miglierati. Optimal weighted least-squares methods. *Preprint*, 2016.

[25] P. G. Constantine. *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*. SIAM, 2015.

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

[27] N. Dyn and M. Floater. Multivariate polynomial interpolation on lower sets. *J. Approx. Theory*, 177:34–42, 2013.

[28] M. I. Ganzburg. Polynomial inequalities on measurable sets and their applications. *Constr. Approx.*, 17:275–306, 2001.

[29] W. Hackbusch. $L^\infty$ estimation of tensor truncations. *Numer. Math.*, 125:419–440, 2013.

[30] M. Lyon and O. Bruno. High-order unconditionally stable FC-AD solvers for general smooth domains II. Elliptic, parabolic and hyperbolic PDEs; theoretical considerations. *J. Comput. Phys.*, 229(9):3358–3381, 2010.
[31] 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.

[32] 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.

[33] 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.

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

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

[36] R. Pasquetti and M. Elghaoui. A spectral embedding method applied to the advection–diffusion equation. *J. Comput. Phys.*, 125:464–476, 1996.

[37] J. Peng, J. Hampton, and A. Doostan. A weighted $\ell_1$-minimization approach for sparse polynomial chaos expansions. *J. Comput. Phys.*, 267:92–111, 2014.

[38] R. Platte, L. N. Trefethen, and A. Kuijlaars. Impossibility of fast stable approximation of analytic functions from equispaced samples. *SIAM Rev.*, 53(2):308–318, 2011.

[39] K. Sargsyan, C. Safta, B. J. Debusschere, and H. Najm. Uncertainty quantification given discontinuous model response and a limited number of model runs. *SIAM J. Sci. Comput.*, 34(1):B44–B64, 2012.

[40] K. Sargsyan, C. Safta, H. Najm, B. J. Debusschere, D. Ricciuto, and P. Thornton. Dimensionality reduction for complex models via Bayesian compressive sensing. *Int. J. Uncertain. Quantif.*, 4(1):63–93, 2014.

[41] K. Stinson, D. F. Gleich, and P. G. Constantine. A randomized algorithm for enumerating zonotope vertices. *arXiv:1602.06620*, 2016.

[42] V. Temlyakov and S. Tikhonov. Remez-type inequalities for the hyperbolic cross polynomials. *arXiv:1606.03773*, 2016.

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

[44] J. A. S. Witteveen and G. Iaccarino. Simplex stochastic collocation with random sampling and extrapolation for nonhypercube probability spaces. *SIAM J. Sci. Comput.*, 34(2):A814–A838, 2012.

[45] D. Xiu and G. E. Karniadakis. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, 24(2):619–644, 2002.

[46] L. Yan, L. Guo, and D. Xiu. Stochastic collocation algorithms using $\ell_1$-minimization. *Int. J. Uncertain. Quantif.*, 2(3):279–293, 2012.

[47] X. Yang and G. E. Karniadakis. Reweighted $\ell_1$ minimization method for stochastic elliptic differential equations. *J. Comput. Phys.*, 248:87–108, 2013.

[48] T. Zhou, A. Narayan, and D. Xiu. Weighted discrete least-squares polynomial approximation using randomized quadratures. *J. Comput. Phys.*, 1:787–800, 2015.