Frames and numerical approximation

BEN ADCOCK* AND DAAN HUYBRECHS†

Abstract. Functions of one or more variables are usually approximated with a basis: a complete, linearly-independent system of functions that spans a suitable function space. The topic of this paper is the numerical approximation of functions using the more general notion of frames: that is, complete systems that are generally redundant but provide infinite representations with bounded coefficients. While frames are well-known in image and signal processing, coding theory and other areas of applied mathematics, their use in numerical analysis is far less widespread. Yet, as we show via a series of examples, frames are more flexible than bases, and can be constructed easily in a range of problems where finding orthonormal bases with desirable properties (rapid convergence, high resolution power, etc.) is difficult or impossible. For instance, we exhibit a frame which yields simple, high-order approximations of smooth, multivariate functions in arbitrary geometries.

A key concern when using frames is that computing a best approximation requires solving an ill-conditioned linear system. Nonetheless, we construct a frame approximation via regularization with bounded condition number (with respect to perturbations in the data), and which approximates any function up to an error of order $\sqrt{\epsilon}$, or even of order ϵ with suitable modifications. Here ϵ is a threshold value that can be chosen by the user. Crucially, rate of decay of the error down to this level is determined by the existence of approximate representations of f in the frame possessing small-norm coefficients. We demonstrate the existence of such representations in all of our examples. Overall, our analysis suggests that frames are a natural generalization of bases in which to develop numerical approximation. In particular, even in the presence of severely ill-conditioned linear systems, the frame condition imposes sufficient mathematical structure in order to give rise to accurate, well-conditioned approximations.

Key words. frames, function approximation, ill-conditioning, singular value decomposition

AMS subject classifications. 42C15, 42C30, 41A10, 65T40

1. Introduction. Frames are a generalization of bases that allow for redundancy amongst the generating elements. They are indispensable tools in modern signal and image processing, and are widely used in a range of other problems, such as compression, source coding, robust transmission and sampling theory [12, 32, 57, 58, 68]. Yet frames – specifically, frames of functions – are generally less well known in numerical analysis.[1] Although they arise in approximation problems in a number of ways, a systematic and general study of numerical frame approximation does not appear to have been undertaken.

The purpose of this paper is to consider frames from this perspective. By means of motivation, we introduce three classes of problems in numerical computing where frames already occur naturally, or where they may potentially lead to better methods. Our main objective is to examine the accuracy and conditioning of frame approximations, and the properties of a given frame which affect both.

A key theme of this paper is the difference between the behaviour of infinite frames and the corresponding truncated frames used in approximation. Quite unlike the case of orthonormal bases, the matrices of the linear systems associated with best approximations in frames (truncated Gram matrices) have condition numbers that necessarily grow with the degree of the approximation. Moreover, this growth

---

*Department of Mathematics, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada (ben_adcock@sfu.ca, http://www.benadcock.ca)

† Department of Computer Science, University of Leuven, Celestijnenlaan 200A, BE-3001 Leuven, Belgium (daan.huybrechs@cs.kuleuven.be, http://people.cs.kuleuven.be/~daan.huybrechs/)

[1]Finite frames of vectors, on the other hand, are better known in the numerical linear algebra community (see, for instance, [36, 44, 74, 82, 83] and references therein). However, the focus of this article is the infinite setting of frames of function spaces.

can be arbitrarily fast. The surprising fact, however, is that accurate frame approximations can still be computed, for instance via suitable regularization. Crucially, and unlike in the case of orthonormal bases, this means that there are fundamental differences between 'theoretical' frame approximations (e.g. the best approximation) and 'numerical' frame approximations (i.e. the solution of the regularized problem). Understanding and documenting these differences is a central aspect of this paper.

1.1. Orthonormal bases. Let $\Phi = \{\phi_n\}_{n\in\mathbb{N}}$ be an orthonormal basis of a Hilbert space H, indexed for convenience over the natural numbers $\mathbb{N} = \{1, 2, \ldots\}$. Two key properties of Φ are as follows. First, the straightforward representation of any $f \in$ H in the basis using the inner product on H. That is,

$$(1) \qquad f = \sum_{n\in\mathbb{N}} \langle f, \phi_n \rangle \phi_n, \qquad \forall f \in \text{H},$$

where the infinite sum converges in H. Note that the *coefficients* $\{\langle f, \phi_n \rangle\}_{n\in\mathbb{N}}$ in this expansion are unique. Second, Parseval's identity, which asserts that the Hilbert space norm of f is precisely the ℓ^2-norm of its coefficients:

$$(2) \qquad \|f\|^2 = \sum_{n\in\mathbb{N}} |\langle f, \phi_n \rangle|^2, \qquad \forall f \in \text{H}.$$

If the coefficients $\{\langle f, \phi_n \rangle\}_{n\in\mathbb{N}}$ are known (or have been computed), approximation in Φ is a straightforward affair. One simply replaces (1) by a finite expansion

$$(3) \qquad f \approx \sum_{n=1}^{N} \langle f, \phi_n \rangle \phi_n.$$

This approximation has the beneficial property of being the orthogonal projection onto the finite-dimensional subspace $\text{H}_N = \text{span}\{\phi_n\}_{n=1}^{N}$, and therefore the best approximation to f from H_N in the norm of H (we assume throughout that H_N is endowed with this norm).

1.2. Frames. An indexed family $\Phi = \{\phi_n\}_{n\in\mathbb{N}}$ is called a *frame* for H if it satisfies the so-called *frame condition*

$$(4) \qquad A\|f\|^2 \leq \sum_{n\in\mathbb{N}} |\langle f, \phi_n \rangle|^2 \leq B\|f\|^2, \quad \forall f \in \text{H},$$

for constants $A, B > 0$. The optimal constants $A, B > 0$ such that (4) holds, i.e. the largest possible A and the smallest possible B, are referred to as the *frame bounds* [32, 41]. We recall frames and their theory in more detail in §2.

Generalizing Parseval's identity (2), the frame condition expresses a norm equivalence between the ℓ^2-norm of the coefficients $\{\langle f, \phi_n \rangle\}_{n\in\mathbb{N}}$ and the Hilbert space norm of f. Yet frames differ from orthonormal bases in a number of key ways:

(i) The frame elements ϕ_n are not generally orthogonal.
(ii) While (4) implies that span(Φ) is dense in H, Φ need not be a basis. In fact, a frame is typically *redundant*. That is, for any $f \in$ H it is possible to find more than one sequence of square-summable coefficients $\boldsymbol{c} = \{c_n\}_{n\in\mathbb{N}}$ such that $f = \sum_{n\in\mathbb{N}} c_n \phi_n$ (with convergence in H).
(iii) In general, a representation such as (1) does not hold.

1.3. Computing orthogonal projections with frames. While (i) means that frames are more flexible than orthonormal bases (indeed, (4) is far less restrictive a condition than orthogonality), it presents an immediate difficulty for numerical approximation with frames. Even if the coefficients $\langle f, \phi_n \rangle$ are known, the orthogonal projection cannot be expressed as in (3) as a sum of these coefficients multiplied by the frame elements. Determining the coefficients $\boldsymbol{x} = \{x_n\}_{n=1}^N$ of the orthogonal projection with respect to the frame elements requires solving a particular linear system (see §2.5 for further details). In the notation of our paper, this is denoted by

$$(5) \qquad \boldsymbol{G}_N \boldsymbol{x} = \boldsymbol{y}, \qquad \boldsymbol{y} = \{\langle f, \phi_n \rangle\}_{n=1}^N,$$

where $\boldsymbol{G}_N$ is the $N \times N$ *truncated Gram matrix*

$$\boldsymbol{G}_N = \{\langle \phi_m, \phi_n \rangle\}_{n,m=1}^N \in \mathbb{C}^{N \times N}.$$

If $\boldsymbol{x} \in \mathbb{C}^N$ is a solution of (5), the orthogonal projection is given by $\sum_{n=1}^N x_n \phi_n$.

The second observation (ii) raises another, and arguably more crucial, issue in practical frame approximation. Due to orthogonality, the Gram matrix of a finite subset $\{\phi_n\}_{n=1}^N$ of an orthonormal basis is perfectly conditioned; indeed, $\boldsymbol{G}_N = \boldsymbol{I}$ is the identity matrix. It is tempting to think that the frame condition (4) endows the Gram matrices of finite subsets of frames with a similar property. However, while the *truncated* frame $\Phi_N = \{\phi_n\}_{n=1}^N$ is indeed a frame for its span H_N, and thus satisfies a frame condition, its frame bounds A_N and B_N may behave wildly as $N \to \infty$, even when the infinite frame bounds A and B are mild. We shall see examples later in this paper where the ratio B_N/A_N grows exponentially fast in N. If Φ_N is linearly independent (a condition which is satisfied by all examples of this paper – see §2.2), the condition number of the Gram matrix $\boldsymbol{G}_N$ is precisely B_N/A_N (Lemma 4). Hence ill-conditioning is equivalent to poorly-behaved frame bounds. Understanding this ill-conditioning and its effect on the resulting numerical frame approximation obtained by solving a regularized version of (5) is the central theme of this paper.

We note in passing that this ill-conditioning stems from the noninvertibility of the Gram *operator* $\mathcal{G}$ of the frame Φ (see §2.2). Such noninvertibility is due to (ii). Discretizing $\mathcal{G}$ with the matrix $\boldsymbol{G}_N$ results in small, nonzero eigenvalues approximating the zero eigenvalue of $\mathcal{G}$ [51], and therefore large condition numbers (see Remark 7).

1.4. Motivations. Orthonormal bases are ubiquitous in numerical analysis. Important cases include Fourier and Chebyshev bases, in which case fast algorithms exist to (approximately) compute the expansions [85]. A major disadvantage of orthogonal bases however is their inflexibility. As an example, consider the problem of approximating smooth functions of one or more variables. While it is easy to construct good[2] orthogonal bases of functions on intervals, it is much harder to do so in higher dimensions unless the domain is particularly simple (e.g. a hypercube). This aside, it is also problematic to find a good basis for singular functions, or to force periodicity on nonperiodic problems in order to take advantage of the FFT.

In this paper we show that good frames can be found for all these problems. In particular, we identify a simple frame with spectral rates of convergence (in the corresponding Hilbert space norm) for approximating functions defined on arbitrary Lipschitz domains; see Fig. 1 below. These examples illustrate three different generic

---

[2]The word 'good' in this paper is taken to mean *spectrally* convergent, i.e. having rates of convergence depending only on the smoothness of the function being approximated.

constructions which always lead to frames: restrictions of orthonormal bases to sub-domains, augmentation of an orthonormal basis by a finite number of additional terms, and concatenation of several orthonormal bases. This leads us to opine that frames are natural tools for many problems in numerical analysis where constructing orthonormal bases is difficult or impossible.

1.5. Overview and main results. We restrict our focus in this paper to two key properties: the convergence of frame approximations and their conditioning. We shall mostly ignore the question of efficiency, since this is highly dependent on the type of frame used and in this paper we strive for generality (we return to this topic briefly in §7). Our main conclusion is the following. In spite of the extreme ill-conditioning of the linear system (5), accurate frame approximations, in a sense we make precise below, can be computed numerically. To do so, the linear system (5) is regularized using a truncated Singular Value Decomposition (SVD) of the Gram matrix $\boldsymbol{G}_N$.

We recall the main elements of frame theory in §2. In §3 we introduce three generic constructions of frames that are useful in numerical approximations along with our three main examples. All these examples deal with the problem of approximating functions where it is not straightforward or even desirable to use orthonormal bases, and where frame approximations present a viable alternative.

Our analysis commences in §4. The ratio B/A of the frame bounds is an important quantity in the conditioning of frames. However, unlike for orthogonal or Riesz bases, passing from the countable frame Φ to a finite subset $\Phi_N = \{\phi_n\}_{n=1}^N$ necessarily causes a deterioration in the frame bounds. We document this phenomenon in §4. Lemma 5 shows how frame bounds deteriorate after truncation, and Proposition 6 establishes that the effect can be arbitrarily bad. The condition numbers $\kappa(\boldsymbol{G}_N)$ for the three example frames are estimated in §4.3.

We consider the computation of the best approximation via orthogonal projection in §5. We first show in Proposition 11 that the ℓ^2-norm of the *exact* solution $\boldsymbol{x}$ of the system (5) is generally unbounded in N, due to the ill-conditioning. Hence, computing the best approximation with any accuracy for large N in floating point arithmetic is typically impossible. However, the situation improves markedly after regularizing $\boldsymbol{G}_N$ by truncating its singular values below a threshold ϵ. In Theorem 13 we show that the convergence of the resulting regularized projection to a function f is dictated by how well f can be approximated by vectors of coefficients with small norm. Specifically, if $\mathcal{P}_N^\epsilon f$ is the regularized projection, then

$$(6) \qquad \|f - \mathcal{P}_N^\epsilon f\| \leq \inf \left\{ \|f - \mathcal{T}_N \boldsymbol{z}\| + \sqrt{\epsilon}\|\boldsymbol{z}\| : \boldsymbol{z} \in \mathbb{C}^N \right\},$$

where $\mathcal{T}_N \boldsymbol{z} = \sum_{n=1}^N z_n \phi_n \in \mathrm{H}_N$. The first term $\|f - \mathcal{T}_N f\|$ is standard and represents the approximation error corresponding to a coefficient vector $\boldsymbol{z}$. The second term $\sqrt{\epsilon}\|\boldsymbol{z}\|$ is uncommon in the literature on frames, and indeed it is specific to a numerical frame approximation. Its effect is to limit the overall approximation error $\|f - \mathcal{T}_N \boldsymbol{z}\| + \sqrt{\epsilon}\|\boldsymbol{z}\|$ by penalizing larger-norm coefficient vectors $\boldsymbol{z}$. In other words, the overall approximation error will only be small asymptotically if f can be represented in the frame (first term) with coefficients of small norm (second term). The existence of such representations in the first place is guaranteed by the frame property, and that is why we argue that the mathematical structure of a frame seems a highly appropriate general context in which to discuss function approximation in redundant systems.

Theorem 14 shows that the coefficient vector $\boldsymbol{x}^\epsilon$ of the regularized projection $\mathcal{P}_N^\epsilon f$ eventually (for large N) exhibits small norm, though there may be an initial regime

in which it is large. The precise result is as follows:

$$(7) \qquad \|\boldsymbol{x}^\epsilon\| \leq \inf \left\{ 1/\sqrt{\epsilon} \|f - \mathcal{T}_N \boldsymbol{z}\| + \|\boldsymbol{z}\| : \boldsymbol{z} \in \mathbb{C}^N \right\}.$$

Again, the frame property is crucial here, in that it ensures the existence of vectors of coefficients $\boldsymbol{z}$ with small norm such that $\mathcal{T}_N \boldsymbol{z} \approx f$. The price to pay for this beneficial property is that the true convergence rate of the best approximation may not be realized after regularization. Instead, one finds best approximations subject to having a small-norm coefficient vector. In practice, however, small-norm coefficient vectors are often more desirable.

A point of clarification. At this stage, the reader may be tempted to conclude that frame approximations are of limited use in practice, since they can obtain at best $\mathcal{O}\left(\sqrt{\epsilon}\right)$ accuracy. Furthermore, Theorem 15 illustrates that the mapping $\boldsymbol{y} \mapsto \mathcal{P}_N^\epsilon f$ from the data $\boldsymbol{y}$ (see (5)) to the approximation $\mathcal{P}_N^\epsilon f$ is not completely well-conditioned, since its (absolute) condition number can be behave like $\mathcal{O}(1/\sqrt{\epsilon})$. However, these issues can be simultaneously overcome. In §6 we briefly describe a generalized frame approximation which achieves $\mathcal{O}\left(\epsilon\right)$ accuracy, and has a bounded condition number, independent of ϵ. The full analysis of these techniques (which builds on this paper) is described in a companion paper [7].

Finally, let us briefly demonstrate the effectiveness of numerical frame approximations. Fig. 1 shows the approximation of ice elevation at the north pole from discrete height data. The domain is highly irregular, and data is also missing at high latitudes. Yet an accurate approximation is achieved using a simple and spectrally-accurate frame, whose elements ϕ_n are nothing more than the usual Fourier basis functions on the bounding box $[-1, 1]^2$ (Example 1 of this paper).

1.6. Relation to existing work. Frames were introduced in the context of nonharmonic Fourier series by Duffin and Schaeffer [41]. They were later developed in the 1980s by Daubechies, Grossmann and Meyer [38] with the systematic study of wavelets. Since then they have become an integral part of modern signal and image processing, compression, coding theory and sampling theory. For overviews, see [12, 22, 26, 32, 37, 57, 58].

With the notable exceptions of [51, 81], [77, 78] and [35, 80], approaches to numerical frame approximations have usually centred around either explicitly identifying an appropriate *dual* frame or numerically inverting the frame *operator* (from which the so-called *canonical dual* frame can be computed) [23, 24, 25, 29, 30, 31, 33, 34]. We refer to §2.3 for further details. While these approaches are useful for approximations with, for example, Gabor or wavelet frames and their various generalizations (e.g. multi-wavelets [32], ridgelets [19], curvelets [20, 68] and shearlets [60]), for the problems which motivate this paper the dual frame expansion usually converges too slowly to be of practical use. We give several examples of this phenomenon later. Conversely, our focus in this paper is on computing best approximations with frames, or more precisely, surrogates obtained from solving regularized systems. The regularized Gram systems we consider in this paper have previously been studied in [51, 81] in the context of frames of exponentials arising in nonuniform sampling problems. In particular, [51, Thm. 5.17] asserts convergence of the coefficients $\boldsymbol{x}^\epsilon$ to the so-called frame coefficients (see §2.3) as $N \rightarrow \infty$. We extend this work by establishing the convergence rate (6) and coefficient bound (7) for arbitrary frames. Note that our regularization procedure is also related to a recent approach of [40], used for the construction of adaptive Gabor frame approximations.

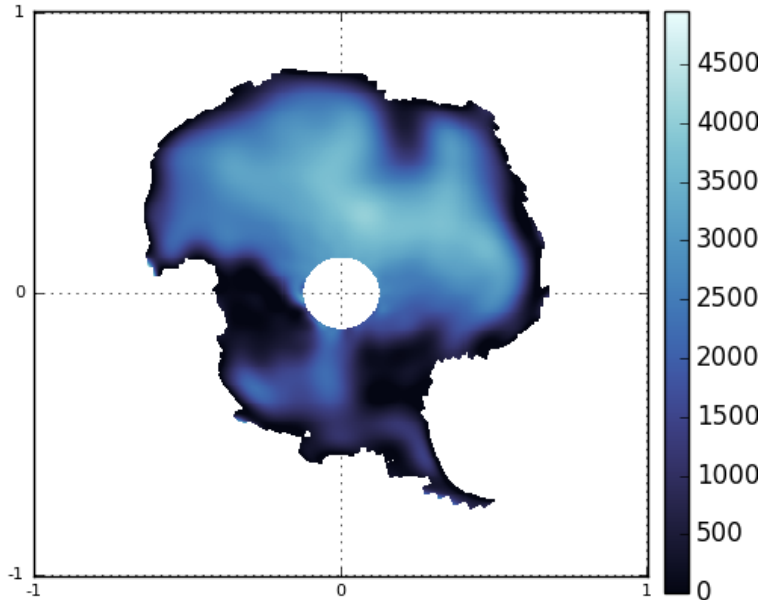Our study of numerical frame approximation stems from previous works on so-

FIG. 1. *Illustration of the elevation of ice at the north pole (in meters). Discrete height data from EOS, NASA's Earth Observing System, was approximated in a discrete least-squares sense using using a conventional two-dimensional Fourier series on a bounding box (the data is publicly available on NASA's EOS website at https://neo.sci.gsfc.nasa.gov/). The data was supplied in equispaced points in polar coordinates, and the approximation was constructed in polar coordinates as well. No data was supplied outside the north pole, i.e. the function samples implicitly define the shape of the domain. Data was also absent beyond a certain latitude, hence the circle of missing data near the pole. The irregular and punctured shape of the domain, without explicitly defined boundary, does not hinder the numerical frame approximation. The truncation parameter ϵ was chosen to be fairly large, in order to reduce the growth of the coefficients while maintaining proximity to the data, as suggested by Theorems 13 and 14. The algorithm used is described and analyzed in [70]. It exploits the equispaced property of the data points via the FFT, even though data is only given on an irregular subset of the full rectangular grid. In this paper, we focus not on the algorithm, but on the mathematical analysis that make this type of approximation possible and reliable.*

called *Fourier extensions* [16, 18], also known as *Fourier continuation* or *Fourier embedding* in the context of numerical PDEs [72]. The connection to frame theory was first explored by the second author in [54], and further developed in the one-dimensional setting by both authors in [8]. A by-product of this paper is an extension of the results of [8] to $d \geq 1$ dimensions. Yet we stress that our main results apply to any frame, not just Fourier extensions.

We also draw several interesting connections to other fields. In particular, the disparity between truncated frames and infinite frames is related to the spectral theory of self-adjoint operators, and specifically the phenomenon of pollution in the finite section method [39, 62]. We also make links to the topic of time- and band-limiting [53], in particular the prolate spheriodal wavefunctions [76], and to classical regularization theory of ill-posed problems [43, 50, 71].

1.7. On supplementary materials. This paper is accompanied by supplementary materials. These materials contain the proofs of a number of results pertaining to the various frames used as examples in the paper. Conversely, all results that are general (that is, apply to any frame) are proved in the main manuscript.

2. Preliminaries.

2.1. Orthogonal and Riesz bases. For the remainder of the paper, $\Phi = \{\phi_n\}_{n \in I}$ is an indexed family in a separable Hilbert space H over the field $\mathbb{C}$, where I is a countable index set (for instance, the natural numbers $\mathbb{N}$ or integers $\mathbb{Z}$). We write $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ for the inner product and norm on H respectively.

The indexed family Φ is an *orthonormal* basis for H if span(Φ) (the vector space of all finite linear combinations of elements of Φ) is dense in H and $\langle \phi_n, \phi_m \rangle = \delta_{n,m}, \forall n, m \in I$. Recall that orthonormal bases satisfy Parseval's identity

$$(8) \qquad \|f\|^2 = \sum_{n \in I} |\langle f, \phi_n \rangle|^2, \quad \forall f \in H.$$

Equivalently,

$$(9) \qquad \|\boldsymbol{x}\| = \|\sum_{n \in I} x_n \phi_n\|, \quad \forall \boldsymbol{x} = \{x_n\}_{n \in I} \in \ell^2(I).$$

Here and throughout, $\ell^2(I)$ denotes the space of square-summable sequences indexed over I, and $\|\cdot\|$ denotes its norm, i.e. $\|\boldsymbol{x}\| = \sqrt{\sum_{n \in I} |x_n|^2}$.

Riesz bases are generalizations of orthonormal bases. Φ is a Riesz basis of H if span(Φ) is dense in H and the following relaxed version of (9) holds:

$$(10) \qquad A\|\boldsymbol{x}\|^2 \leq \|\sum_{n \in I} x_n \phi_n\|^2 \leq B\|\boldsymbol{x}\|^2, \quad \forall \boldsymbol{x} = \{x_n\}_{n \in I} \in \ell^2(I).$$

Here $A, B > 0$ are positive constants. Throughout this paper, whenever constants A and B are introduced in an inequality such as this, they will be taken to be the optimal constants such that the inequality holds. Note that this inequality also implies the following relaxed version of (8):

$$(11) \qquad A\|f\|^2 \leq \sum_{n \in I} |\langle f, \phi_n \rangle|^2 \leq B\|f\|^2, \quad \forall f \in H.$$

Every orthonormal basis is a Riesz basis, but Riesz bases need not be orthonormal, or even orthogonal. A popular example of such a basis are the hat functions commonly used in finite element methods. They are not orthogonal, but they are a basis for their span. Another example are more general B-splines [32].

Any Riesz basis has a unique *dual* Riesz basis Ψ. This basis satisfies

$$\langle \phi_n, \psi_m \rangle = \delta_{n,m}, \quad n, m \in I.$$

For this reason, the Riesz basis and its dual are termed *biorthogonal*. From this, the unique representation of an element $f \in H$ in the Riesz basis is given explicitly in terms of inner products with the dual basis. That is,

$$(12) \qquad f = \sum_{n \in I} \langle f, \psi_n \rangle \phi_n = \sum_{n \in I} \langle f, \phi_n \rangle \psi_n,$$

with convergence in H. Note that an orthonormal basis is self-dual, i.e. $\Psi = \Phi$.

2.2. Frames. Φ is called a *frame* for H if it satisfies the *frame condition*, which we repeat here for convenience:

$$(13) \qquad A\|f\|^2 \leq \sum_{n \in I} |\langle f, \phi_n \rangle|^2 \leq B\|f\|^2, \quad \forall f \in \mathrm{H}.$$

Note that (13) implies that span(Φ) is dense in H. It follows from the Parseval identity (8) and its generalization (11) that orthonormal and Riesz bases are frames. However, most frames are not bases. Indeed, frames are generally not ω-*independent* (see [32, Sec. 6.1], for example): that is, there exist nonzero coefficients $\{x_n\}_{n \in I}$ for which the sum $\sum_{n \in I} x_n \phi_n$ converges in H and satisfies $\sum_{n \in I} x_n \phi_n = 0$. Conversely, bases are always ω-independent. As mentioned, this *redundancy* gives frames far greater flexibility than bases, making them easier to construct for particular problems.

We now introduce several standard concepts in frame theory. Associated to any frame Φ is the so-called *synthesis* operator

$$\mathcal{T} : \ell^2(I) \to \mathrm{H}, \quad \boldsymbol{y} = \{y_n\}_{n \in I} \mapsto \sum_{n \in I} y_n \phi_n.$$

Its adjoint, the *analysis* operator, is given by

$$\mathcal{T}^* : \mathrm{H} \to \ell^2(I), \quad f \mapsto \{\langle f, \phi_n \rangle\}_{n \in I},$$

and the composition $\mathcal{S} = \mathcal{T}\mathcal{T}^*$, known as the *frame* operator, is

$$\mathcal{S} : \mathrm{H} \to \mathrm{H}, \quad f \mapsto \sum_{n \in I} \langle f, \phi_n \rangle \phi_n.$$

Note $\mathcal{S}f$ converges to f in H for an orthonormal basis, but for a Riesz basis or frame this is no longer the case in general. Still, the frame operator is a useful object. It is self-adjoint by construction, and it follows from the frame condition that $\mathcal{S}$ is also bounded and invertible on H [32, Lemma 5.1.5].

The Gram operator of a frame is defined by $\mathcal{G} = \mathcal{T}^*\mathcal{T}$. That is,

$$(14) \qquad \mathcal{G} : \ell^2(I) \to \ell^2(I), \ \boldsymbol{x} = \{x_n\}_{n \in I} \mapsto \left\{ \sum_{m \in I} \langle \phi_m, \phi_n \rangle x_m \right\}_{n \in I}.$$

Note that $\mathcal{G}$ is a bounded operator on $\ell^2(I)$, but is not in general invertible (see §4). We may also view $\mathcal{G}$ as the infinite matrix $\boldsymbol{G} = \{\langle \phi_n, \phi_m \rangle\}_{n,m \in I}$. Throughout this paper all infinite matrices are equivalent to bounded operators on $\ell^2(I)$.

A frame is said to be *tight* if $A = B$, in which case $\mathcal{S} = A\mathcal{I}$ is a multiple of the identity $\mathcal{I}$. However, the corresponding Gram operator $\mathcal{G}$ of a tight frame is not a multiple of the identity, unless the frame happens to be an orthonormal basis (for its closed linear span).

We shall also need two further notions. First, a frame is said to be *exact* if it ceases to be a frame when any one element is removed. A frame that is not exact is referred to as *inexact*. Second, we say a frame $\{\phi_n\}_{n \in I}$ is *linearly independent* if every finite subset $\{\phi_n\}_{n \in J}$, $|J| < \infty$, is linearly independent.

A frame is exact if and only if it is a Riesz basis [32, Theorem 5.5.4]. Hence, for the remainder of this paper we will assume that all frames are inexact. We shall also assume that all frames are linearly independent. This is mainly for convenience, and

it will be the case in all examples discussed. Note that a linearly-independent frame is not necessarily a Riesz basis. See [32, Chpt. 6] for further discussion on independence and the relations between frames and Riesz bases.

2.3. Dual frames. A frame $\Psi = \{\psi_n\}_{n\in I} \subseteq \mathrm{H}$ is called a *dual* frame for Φ if

$$(15) \qquad f = \sum_{n\in I}\langle f, \psi_n\rangle\phi_n = \sum_{n\in I}\langle f, \phi_n\rangle\psi_n, \qquad \forall f \in \mathrm{H}.$$

An inexact frame necessarily has more than one dual frame. Moreover, a frame and its duals are not biorthogonal, unlike in the case of Riesz bases. However, there is a unique so-called *canonical* dual frame $\Psi = \{\psi_n\}_{n\in I}$, with elements given by $\psi_n = \mathcal{S}^{-1}\phi_n$ where $\mathcal{S}$ is the frame operator. Since Ψ is a dual frame, one has

$$(16) \qquad f = \sum_{n\in I}\langle f, \mathcal{S}^{-1}\phi_n\rangle\phi_n = \sum_{n\in I}\langle \mathcal{S}^{-1}f, \phi_n\rangle\phi_n.$$

Furthermore, the canonical dual frame bounds are $1/B$ and $1/A$ respectively, i.e.

$$(17) \qquad 1/B\|f\|^2 \le \sum_{n\in I}|\langle f, \mathcal{S}^{-1}\phi_n\rangle|^2 \le 1/A\|f\|^2, \quad \forall f \in \mathrm{H}.$$

We refer to the coefficients $\boldsymbol{a} = \{\langle f, \mathcal{S}^{-1}\phi_n\rangle\}_{n\in I}$ as the *frame coefficients* of f. Note that these coefficients have the beneficial property that, amongst all possible representations of f in Φ, they have the smallest norm. That is, if $f = \sum_{n\in I} a_n\phi_n = \sum_{n\in I} c_n\phi_n$ for some $\boldsymbol{c} = \{c_n\}_{n\in I}$, then $\|\boldsymbol{c}\| \ge \|\boldsymbol{a}\|$ [32, Lem. 5.4.2].

At this stage, one might be tempted to approximate f by computing its dual frame coefficients and truncating the expansion (15). As noted in §1.6, this could potentially be done either by analytically identifying a dual frame (when possible) or by numerically inverting the frame operator. However, computational issues aside – in the case where the frame is not tight, this requires inversion of the operator $\mathcal{S}$, which has infinite-dimensional domain and range – the approximation $\sum_{n\in I_N}\langle \mathcal{S}^{-1}f, \phi_n\rangle\phi_n$ is generally not the orthogonal projection onto $\mathrm{H}_N = \mathrm{span}\{\phi_n : n \in I_N\}$. For the examples which motivate this paper, this approximation typically converges much more slowly in the norm of H than the orthogonal projection.[3] See §3 and Figs. 4 and 5 for several examples of this phenomenon.

2.4. Truncated frames. For each $N \in \mathbb{N}$ we introduce the truncated system $\Phi_N = \{\phi_n\}_{n\in I_N}$ where $I_N \subseteq I$ is a finite index set with $|I_N| = N$. For simplicity, we assume that the index sets $\{I_N\}_{N\in\mathbb{N}}$ are nested and satisfy

$$(18) \qquad I_1 \subseteq I_2 \subseteq \ldots, \qquad \bigcup_{N=1}^{\infty} I_N = I.$$

The system Φ_N is a frame for its span $\mathrm{H}_N = \mathrm{span}(\Phi_N)$. We write $A_N, B_N > 0$ for the frame bounds, so that

$$(19) \qquad A_N\|f\|^2 \le \sum_{n\in I_N}|\langle f, \phi_n\rangle|^2 \le B_N\|f\|^2, \quad \forall f \in \mathrm{H}_N,$$

---

[3]This is in contrast to the case of wavelet frames and their various generalizations, which are specifically designed to have accurate dual frame representations [32, 68].

and let

$$\mathcal{T}_N : \mathbb{C}^N \to \mathrm{H}_N, \ \boldsymbol{y} = \{y_n\}_{n \in I_N} \mapsto \sum_{n \in I_N} y_n \phi_n,$$

(20)
$$\mathcal{T}_N^* : \mathrm{H}_N \to \mathbb{C}^N, \ f \mapsto \{\langle f, \phi_n \rangle\}_{n \in I_N},$$

$$\mathcal{S}_N = \mathcal{T}_N \mathcal{T}_N^* : \mathrm{H}_N \to \mathrm{H}_N, \ f \mapsto \sum_{n \in I_N} \langle f, \phi_n \rangle \phi_n,$$

be the truncated synthesis, analysis and frame operators respectively. We also define the truncated Gram operator $\mathcal{G}_N = \mathcal{T}_N^* \mathcal{T}_N$ and the associated $N \times N$ Gram matrix

(21)
$$\boldsymbol{G}_N = \{\langle \phi_m, \phi_n \rangle\}_{n,m \in I_N} \in \mathbb{C}^{N \times N}.$$

Since the frame Φ is linearly independent by assumption, it follows that the Gram matrix $\boldsymbol{G}_N$ is nonsingular. Indeed, if $\boldsymbol{x} = \{x_n\}_{n \in I_N} \in \mathbb{C}^N$ then $\boldsymbol{x}^* \boldsymbol{G}_N \boldsymbol{x} = \left\| \sum_{n \in I_N} x_n \phi_n \right\|^2$ and, by linear independence, the right-hand side is zero if and only if $\boldsymbol{x} = \boldsymbol{0}$.

2.5. Best approximations and rates of convergence. A key task in approximation is to compute the orthogonal projection $\mathcal{P}_N$ onto $H_N = \mathrm{span}(\Phi_N)$. Observe that $\mathcal{P}_N f$ is the best approximation to f from H_N in $\|\cdot\|$. For $f \in \mathrm{H}$, write

(22)
$$\mathcal{P}_N f = \sum_{n \in I_N} x_n \phi_n, \quad \boldsymbol{x} = \{x_n\}_{n \in I_N} \in \mathbb{C}^N.$$

Since $\mathcal{P}_N f$ is defined by the orthogonality conditions

$$\langle \mathcal{P}_N f, \phi_n \rangle = \langle f, \phi_n \rangle, \quad \forall n \in I_N,$$

the coefficients $\boldsymbol{x} = \{x_n\}_{n \in I_N}$ are the unique solution of the linear system

(23)
$$\boldsymbol{G}_N \boldsymbol{x} = \boldsymbol{y}, \quad \boldsymbol{y} = \{\langle f, \phi_n \rangle\}_{n \in I_N}.$$

Hence, computing the best approximation in a frame requires solving an $N \times N$ linear system. This system turns out to be ill-conditioned, in direct contrast with the case of an orthonormal basis, wherein the Gram matrix $\boldsymbol{G}_N$ is the identity and $x_n = y_n = \langle f, \phi_n \rangle$. The treatment of this ill-conditioning is central task of this paper.

Besides conditioning, another concern of this paper is the convergence rate of the approximation $\mathcal{P}_N f$ and its surrogates obtained via regularization. To this end, we distinguish three types of convergence of an approximation f_N to a function f. First, we say that f_N converges *algebraically* fast to f at rate k if $\|f - f_N\| = \mathcal{O}(N^{-k})$ as $N \to \infty$. Second, if $\|f - f_N\|$ decays faster than any algebraic power of N^{-1} then we say that f_N converges *superalgebraically* fast to f. Third, we say that f_N converges *geometrically* fast to f if there exists a $\rho > 1$ such that $\|f - f_N\| = \mathcal{O}(\rho^{-N})$.

3. Examples of frames. We now introduce three examples of frames that will be used throughout the paper to interpret our main results. These examples also serve to illustrate the flexibility gained by allowing redundancy in the approximation.

Example 1. Fourier frames for complex geometries. Let $\Omega \subseteq \mathbb{R}^d$ be a compact domain and $f : \Omega \to \mathbb{R}$. Besides simple domains (cubes, toruses, spheres, etc.), it is in general difficult to find orthonormal bases for $\mathrm{H} = \mathrm{L}^2(\Omega)$ with simple, explicit expressions and whose orthogonal projections converge *spectrally* fast: that is, algebraic

for functions with finite orders of smoothness and superalgebraic for infinitely-smooth functions. However, it is straightforward to find a frame with this property.

Since Ω is compact, it can be contained in an open hypercube. Without loss of generality, suppose that $\Omega \subseteq (-1, 1)^d$. Now consider a system of functions formed by the restriction of the orthonormal Fourier basis on $(-1, 1)^d$ to Ω:

$$(24) \qquad \Phi = \{\phi_{\boldsymbol{n}}\}_{\boldsymbol{n}\in\mathbb{Z}^d}, \qquad \phi_{\boldsymbol{n}}(\boldsymbol{t}) = 2^{-d/2}\mathrm{e}^{\mathrm{i}\pi\boldsymbol{n}\cdot\boldsymbol{t}}, \quad \boldsymbol{t} \in \Omega.$$

This system is not an orthonormal basis of $\mathrm{L}^2(\Omega)$, but it is a tight, linearly-independent frame with $A = B = 1$. If we introduce the truncated frames
$$(25)$$
$$\Phi_N = \{\phi_{\boldsymbol{n}}\}_{\boldsymbol{n}\in I_N}, \quad I_N = \left\{ \boldsymbol{n} = (n_1, \ldots, n_d) \in \mathbb{Z}^d : -\tfrac{1}{2}N^{1/d} \le n_1, \ldots, n_d < \tfrac{1}{2}N^{1/d} \right\},$$

(we assume throughout that $\frac{1}{2}N^{1/d}$ is a positive integer) then the convergence rate of the resulting orthogonal projections $\mathcal{P}_N f$ is spectral (Proposition 17).

The approximation based on the frame (24) is known as a Fourier *extension* (or *continuation*) [16, 18] in the one-dimensional case, and occasionally referred to as a Fourier *embedding* in higher dimensions [15, 72]. The connection to frames was first explored in [54], and further analysis of the $d = 1$ case was given in [6, 8, 65].

Recalling the discussion in §2.3, this frame is an example where the canonical dual frame expansion (16) converges slowly. Indeed, since Φ is tight with $A = B = 1$ it is its own canonical dual frame, and therefore the frame coefficients are $a_{\boldsymbol{n}} = \langle f, \phi_{\boldsymbol{n}}\rangle$. They are precisely the Fourier coefficients of the extension $\tilde{f}$ of f by zero to $(-1, 1)^d$:

$$\langle f, \phi_{\boldsymbol{n}}\rangle = \int_\Omega f(\boldsymbol{t})\phi_{\boldsymbol{n}}(\boldsymbol{t})\mathrm{d}\boldsymbol{t} = \int_{(-1,1)^d} \tilde{f}(\boldsymbol{t})\phi_n(\boldsymbol{t})\mathrm{d}\boldsymbol{t}.$$

As a result, the expansion (16) is nothing more than the Fourier series of $\tilde{f}$ restricted to Ω. Unless f vanishes smoothly on the boundary $\partial\Omega$, this expansion converges slowly and suffers from Gibbs' phenomenon near $\partial\Omega$. In contrast, the convergence of the best approximation $\mathcal{P}_N f$ is spectral, regardless of the shape of Ω.

REMARK 1. *This example illustrates a general principle: the restriction of a Riesz basis on a domain Ω_e to a subset $\Omega \subset \Omega_e$ always results in a frame. If the basis on Ω_e is orthonormal, the corresponding frame on Ω is tight. Such a construction (albeit without the connection to frame theory) has been used for the numerical solution of PDEs in complex geometries. Recent examples of embedding methods implicitly based on such 'extension' frames include [64, 79]. See also [10, 17, 67] for a method based on one-dimensional extensions. Besides the Fourier extensions, other examples of extension schemes in include volume penalty methods [75], immersed domain methods [13], and embedded or fictitious domain methods [56].*

Example 2. Augmented Fourier basis. Consider the case of smooth, nonperiodic functions $f : [-1, 1] \to \mathbb{R}$. Polynomial bases have good convergence properties for such functions but relatively bad resolution power for oscillatory functions. On the other hand, the Fourier approximation of a nonperiodic function suffers from Gibbs' phenomenon at $t = \pm 1$ and converges only slowly in the L^2-norm. One way to seek to remedy this situation is to augment the Fourier basis by a finite number $K \in \mathbb{N}$ of polynomials $\psi_1, \ldots, \psi_K$, leading to the system

$$(26) \qquad \Phi = \{\varphi_n\}_{n\in\mathbb{Z}} \cup \{\psi_k\}_{k=1}^K, \qquad \varphi_n(t) = \tfrac{1}{\sqrt{2}}\mathrm{e}^{\mathrm{i}n\pi t}.$$

To save unnecessary generalizations, we assume that $\psi_k = \sqrt{k+1/2}P_k$, where $P_k \in \mathbb{P}_k$ is the kth Legendre polynomial. Note that $\{\psi_k\}_{k=1}^K$ is an orthonormal basis for

$$\mathbb{P}_K^0 = \left\{ p \in \mathbb{P}_K : \int_{-1}^1 p(t)\,\mathrm{d}t = 0 \right\}.$$

Since $\{\varphi_n\}_{n\in\mathbb{Z}}$ is an orthonormal basis and K is finite, Φ forms a frame for H $=$ $\mathrm{L}^2(-1,1)$ with frame bounds $A = 1$ and $B = 2$. It is also linearly independent, since no finite sum of the complex exponentials φ_n is exactly equal to a nonconstant algebraic polynomial. If

$$\Phi_N = \{\varphi_n : n = -\tfrac{N-K}{2}, \ldots, \tfrac{N-K}{2} - 1\} \cup \{\psi_k\}_{k=1}^K, \quad N \geq K, \quad N - K \text{ even,}$$

is the truncated frame (we will not consider the odd case, although it presents few difficulties), then orthogonal projections with respect to this frame inherit the optimal resolution properties of the Fourier basis, yet converge algebraically at rate K for all sufficiently smooth functions (Proposition 18). Conversely, the canonical dual frame expansion converges at roughly the same rate as the Fourier expansion of f (see Proposition SM5).

The idea of augmenting the Fourier basis with a finite number of additional functions is an old one, arguably dating back to Krylov [59]. These functions endow the frame with good approximation properties by implicitly subtracting the jump discontinuity of f and its first $K-1$ derivatives at $t = \pm 1$. This smoothed function now has faster decaying Fourier coefficients, leading to more rapid convergence of $\mathcal{P}_N f$. This approach is also referred to as Eckhoff's method [42] or Euler–MacLaurin interpolants [55]. Whilst the convergence of this approximation has been extensively studied (see [1, 2] and references therein), the connection with frame theory is, to the best of our knowledge, new.

REMARK 2. *An overall principle illustrated by this example is that adding a finite number of elements of* H *to a Riesz (in particular, orthogonal) basis always results in a frame; a so-called Riesz frame [32, Sec. 6.2]. Note that such a frame is generally not linearly independent. Although we focus on polynomials enhancing the Fourier basis here, augmenting a basis with additional terms to incorporate features of the function to be approximated (in this case, smoothness) is quite a general idea. Other examples include piecewise polynomial functions in the presence of interior discontinuities, or compactly-supported functions in the case of local variations such as oscillations.*

Example 3. Polynomials plus modified polynomials.
Several problems in numerical analysis call for the approximation of functions of the form

$$(27) \qquad\qquad f(t) = w(t)g(t) + h(t), \quad t \in [-1, 1],$$

where g, h are smooth functions but $w \in \mathrm{L}^\infty(-1,1)$ may be singular, oscillatory or possessing some other kind of feature which makes approximation difficult. The presence of $w(t)$ usually means that the polynomial approximation of f converges only slowly in N. One particular instance is

$$(28) \qquad\qquad w(t) = (1+t)^\alpha, \quad 0 < \alpha < 1,$$

which corresponds to a weak endpoint singularity of the function f. Note that (28) has been considered in [27, 28] in the context of PDEs with endpoint singularities. For other examples corresponding to an oscillatory function $w(t)$, see [52, 73].

If $w(t)$ is known or has been approximated (as in the applications mentioned above), then it is natural to use it to construct a frame to approximate f. Let $\{\varphi_n\}_{n \in \mathbb{N}}$ be the orthonormal basis of Legendre polynomials (we could also use Chebyshev polynomials here, but we shall use Legendre for simplicity). Then we form the system

$$(29) \qquad\qquad \Phi = \{\varphi_n\}_{n \in \mathbb{N}} \cup \{\psi_n\}_{n \in \mathbb{N}}, \qquad \psi_n(t) = w(t)\varphi_n(t).$$

Since $w \in \mathrm{L}^\infty(-1,1)$ and $\{\varphi_n\}_{n \in \mathbb{N}}$ is an orthonormal basis, this system gives rise to a frame for the space $\mathrm{H} = \mathrm{L}^2(-1,1)$. The frame bounds are

$$A = 1 + \operatorname*{ess\,inf}_{t \in (-1,1)} |w(t)|^2, \quad B = 1 + \operatorname*{ess\,sup}_{t \in (-1,1)} |w(t)|^2.$$

This frame is linearly dependent if and only if w is a rational function of two polynomials. We assume from now on that this is not the case. For even N, we define the truncated frames by $\Phi_N = \{\varphi_n\}_{n=1}^{N/2} \cup \{\psi_n\}_{n=1}^{N/2}$. Orthogonal projections with respect to this frame are spectrally convergent with respect to the smoothness of g and h (Proposition 19). Conversely, the convergence of the dual frame expansion is generally not spectral, but algebraic at a fixed and low rate (Proposition SM8).

Note that more terms can be included in (27), i.e. $f(t) = \sum_{i=1}^{K} w_i(t)g_i(t)$ for functions $w_1, \dots, w_K$. If these are known, then this would lead to the frame construction $\Phi = \cup_{i=1}^{K} \{\psi_{i,n}\}_{n \in \mathbb{N}}$, where $\psi_{i,n}(t) = w_i(t)\varphi_n(t)$. For simplicity, we consider only the case (27), although the generalization is conceptually straightforward. As in Example 2, the interpretation of this approach as a frame approximation has not, to the best to the best of our knowledge, been considered before.

REMARK 3. *The concatenation of a finite number of Riesz or orthonormal bases always results in a frame (albeit one which is generally not linearly independent). More generally, the composition of several frames is still a frame. We note that the concept of concatenations of bases or frames is widely-used in signal and image processing [21]. Typically, images and signals may have substantially sparser representations in the resulting frame than in a single orthonormal basis, which yields benefits in tasks such as compression and denoising [68].*[4]

4. Truncated Gram matrices and ill-conditioning. Since we have assumed linear independence, the truncated frame Φ_N is also a Riesz basis for its span. Hence the truncated frame bounds A_N and B_N are the same as the Riesz bounds. However, this finite basis is very skewed, and this results in the ill-conditioning of truncated frames. In this section we explore this phenomenon in more detail.

4.1. The Gram operator. The Gram operator (14) of a frame Φ is self-adjoint and nonnegative on $\ell^2(I)$ with closed range. It is bounded, and its restriction $\mathcal{G} : \ell^2(I) \to \mathrm{Ran}(\mathcal{G})$ is invertible. Its spectrum $\sigma(\mathcal{G})$ satisfies

$$\{0\} \subseteq \sigma(\mathcal{G}) \subseteq \{0\} \cup [A, B],$$

where A, B are the frame bounds [51]. As shown in [84], $\mathcal{G}$ is compact if and only if H is finite-dimensional, and $\mathcal{G}$ is positive if and only if Φ is a Riesz basis. Hence,

---

[4]For completeness, we note that the system $\{\psi_n\}_{n \in \mathbb{N}}$ in (29) is only a Riesz basis if $w(t)$ is bounded away from zero. This is not the case in (28). However, $\{\psi_n\}_{n \in \mathbb{N}}$ is always a *Bessel sequence* (see, for example, [32, Def. 3.2.2]); that is, a sequence for which the upper frame condition $\sum_{n \in \mathbb{N}} |\langle f, \psi_n \rangle|^2 \leq B\|f\|^2$ holds, but for which the lower frame condition need not hold. The composition of a Riesz basis or frame with a finite number of Bessel sequences is also always a frame.

in this paper $\mathcal{G}$ is singular and thus $\mathrm{Ker}(\mathcal{G}) \neq \{\mathbf{0}\}$. Nonetheless, since $\mathcal{G}$ has closed range, we may define its Moore–Penrose pseudoinverse $\mathcal{G}^\dagger : \ell^2(I) \to \ell^2(I)$ [84]. One then has the following relation between $\mathcal{G}$, $\mathcal{G}^\dagger$ and the frame bounds:

$$(30) \qquad\qquad A = \left\| \mathcal{G}^\dagger \right\|^{-1}, \quad B = \|\mathcal{G}\|.$$

Here $\|\cdot\|$ is the operator norm on $\ell^2(I)$.

4.2. Truncated Gram matrices. We now consider the condition number of the matrix $\boldsymbol{G}_N$. Following the discussion above note the following standard result:

LEMMA 4. *The truncated Gram matrix $\boldsymbol{G}_N$ of a linearly-independent frame Φ is invertible with $\|\boldsymbol{G}_N^{-1}\|^{-1} = A_N$ and $\|\boldsymbol{G}_N\| = B_N$, where A_N and B_N are the frame bounds of the truncated frame Φ_N. In particular, the condition number of $\boldsymbol{G}_N$ is precisely the ratio of the truncated frame bounds: $\kappa(\boldsymbol{G}_N) = \|\boldsymbol{G}_N\|\|\boldsymbol{G}_N^{-1}\| = B_N/A_N$.*

In practice, we will also use the following characterization of the frame bounds:

$$(31) \qquad\qquad A_N = \min_{\substack{\boldsymbol{x}\in\mathbb{C}^N \\ \|\boldsymbol{x}\|=1}} \|\mathcal{T}_N \boldsymbol{x}\|^2, \quad B_N = \max_{\substack{\boldsymbol{x}\in\mathbb{C}^N \\ \|\boldsymbol{x}\|=1}} \|\mathcal{T}_N \boldsymbol{x}\|^2,$$

which follows from the fact that $\boldsymbol{G}_N = \mathcal{T}_N^* \mathcal{T}_N$. This characterization leads to several different interpretations of the frame bounds. First, A_N measures how small an element of H_N can be while having unit-norm coefficients in the frame. Equivalently, it measures how well the zero element $0 \in \mathrm{H}$ can be approximated by a nonzero element H_N with unit-norm coefficients. It is clear from this that when frame elements are close to being linearly dependent, the constant A_N can be quite small. Conversely, B_N measures how large an element H with unit-norm coefficients can be. Hence, one expects B_N to remain bounded even for near linearly-dependent frame elements.

LEMMA 5. *Let Φ be a linearly-independent frame. Then*
  (i)   the sequences $\{A_N\}_{N\in\mathbb{N}}$ and $\{B_N\}_{N\in\mathbb{N}}$ are monotonically nonincreasing and nondecreasing respectively,
  (ii) $B_N \leq B$ for all N and $B_N \to B$ as $N \to \infty$,
  (iii) $\inf_N A_N > 0$ if and only if Φ is a Riesz basis.

Proof. Part (i) follows immediately from (18) and (31), as does the the observation that $B_N \leq B$ in part (ii). To deduce convergence, let $0 < \epsilon < \sqrt{B}$ be arbitrary and suppose that $\boldsymbol{x} \in \ell^2(I)$, $\|\boldsymbol{x}\| = 1$, is such that $\sqrt{B} \geq \|\mathcal{T}\boldsymbol{x}\| = \sqrt{\langle \mathcal{S}\boldsymbol{x}, \boldsymbol{x}\rangle} \geq \sqrt{B} - \epsilon$. Let $\boldsymbol{z} \in \mathbb{C}^N$ be such that $z_n = x_n$ for $n \in I_N$ and suppose that $\boldsymbol{x}_N \in \ell^2(I)$ is the extension of $\boldsymbol{z}$ by zero. Then

$$\sqrt{B_N} \geq \frac{\|\mathcal{T}_N \boldsymbol{z}\|}{\|\boldsymbol{z}\|} = \frac{\|\mathcal{T}\boldsymbol{x}_N\|}{\|\boldsymbol{x}_N\|} \geq \frac{\|\mathcal{T}\boldsymbol{x}\| - \|\mathcal{T}(\boldsymbol{x} - \boldsymbol{x}_N)\|}{\|\boldsymbol{x}\| + \|\boldsymbol{x} - \boldsymbol{x}_N\|} \geq \frac{\sqrt{B} - \epsilon - \sqrt{B}\|\boldsymbol{x} - \boldsymbol{x}_N\|}{1 + \|\boldsymbol{x} - \boldsymbol{x}_N\|}.$$

Since $\boldsymbol{x}_N \to \boldsymbol{x}$ as $N \to \infty$, we deduce that $\sqrt{B} \geq \sqrt{B_N} \geq \sqrt{B} - 2\epsilon$ for all sufficiently large N. Since ϵ was arbitrary we see that $B_N \to B$. Finally, for part (iii) we use [32, Prop. 6.1.2]. □

These lemmas imply that the truncated Gram matrices $\boldsymbol{G}_N$ are necessarily ill-conditioned for large N. Such ill-conditioning can also be arbitrarily bad:

PROPOSITION 6. *Let $\{\varphi_n\}_{n\in\mathbb{N}}$ be an orthonormal basis of H and let $g \in \mathrm{H}$, $\|g\| = 1$, be such that $\langle g, \varphi_n\rangle \neq 0$ for infinitely many n. Then the system $\Phi = \{g, \varphi_1, \varphi_2, \ldots\}$*
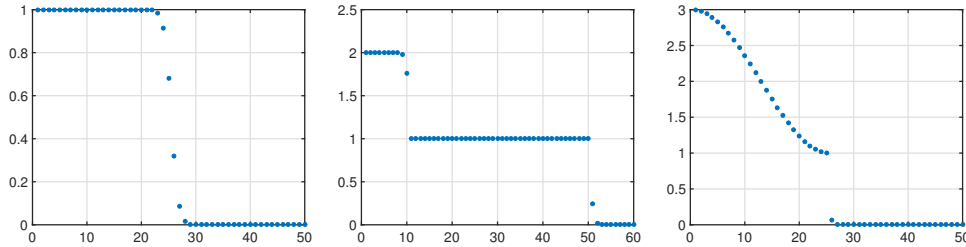
FIG. 2. *The eigenvalues of $\boldsymbol{G}_{50}$ for Examples 1–3 (left to right). The parameters used were $T = 2$ (Example 1), $K = 10$ (Example 2) and $w(t) = \sqrt{1+t}$ (Example 3).*

is a linearly-independent frame for H with bounds $A = 1$ and $B = 2$. Moreover, if $\Phi_N = \{g, \varphi_1, \phi_2, \ldots, \varphi_{N-1}\}$, then the finite frame bounds are given by

$$(32) \qquad A_N = 1 - \sqrt{\sum_{n=1}^{N-1} |\langle g, \varphi_n \rangle|^2}, \qquad B_N = 1 + \sqrt{\sum_{n=1}^{N-1} |\langle g, \varphi_n \rangle|^2}.$$

See §SM1 for the proof. Observe that $A_N \leq \sqrt{\sum_{n \geq N} |\langle g, \varphi_n \rangle|^2}$ since $\|g\| = 1$ for this frame. Hence the behaviour of A_N is related to the decay of the coefficients $\langle g, \varphi_n \rangle$ in the basis $\{\varphi_n\}_{n \in \mathbb{N}}$. The better g is approximated in this basis, the worse the conditioning of the truncated Gram matrix.

This result illustrates how easy it is for the truncated Gram matrices to be ill-conditioned: we can create arbitrarily bad conditioning merely by augmenting an orthonormal basis with one additional element that is well approximated in the basis. When more elements are added (as in Example 2) or a whole orthonormal basis is added (as in Example 3), it is therefore not surprising that the corresponding Gram matrices can be exceedingly ill-conditioned. See §4.3 for further details.

REMARK 7. *The truncated Gram matrices $\boldsymbol{G}_N$ are the $N \times N$ finite sections of the Gram operator $\mathcal{G}$. Recall that $\{0\} \subseteq \sigma(\mathcal{G}) \subseteq \{0\} \cup [A, B]$. Unfortunately, as illustrated in Fig. 2 for the three examples, $\sigma(\boldsymbol{G}_N)$ does not lie within $\{0\} \cup [A, B]$. Instead, spurious eigenvalues are introduced in the spectral gap between 0 and the lower frame bound A; a well-known phenomenon referred to as pollution in the finite section method [39, 62]. The small eigenvalues of $\boldsymbol{G}_N$ are of particular relevance to this paper. These translate into ill-conditioning of $\boldsymbol{G}_N$, meaning that the system (23) requires regularization. Note that $0 \notin \sigma(\boldsymbol{G}_N)$ since the frame is linearly independent, yet as seen in Lemma 5, small eigenvalues necessarily arise as approximations to the zero eigenvalue of $\mathcal{G}$.*

4.3. Examples. We now consider $\kappa(\boldsymbol{G}_N)$ for the examples of §3. Proofs of the results presented in this section can be found in the supplementary material.

Fourier frames for complex geometries. Consider the frame of Example 1. If $\Omega = (-\frac{1}{T}, \frac{1}{T})$ is an interval, where $T > 1$, then $\kappa(\boldsymbol{G}_N) = \mathcal{O}\left(E(T)^N\right)$ as $N \to \infty$, where $E(T) = \cot^2(\pi/(4T)) > 1$ [8]. Hence the condition number is geometrically large in N – see Fig. 3(a). A similar, albeit somewhat weaker, result also holds in arbitrary dimensions:

PROPOSITION 8. *Let $\Omega \subseteq (-1,1)^d$ be a compact domain with Lipschitz boundary and consider the frame* (24). *Then the condition numbers $\kappa(\boldsymbol{G}_N)$ grow superalgebraically fast in N.*

The explanation of this result is rather simple. One can show that the kernel $\mathrm{Ker}(\mathcal{G})$, a subset of $\ell^2(I)$, consists precisely of the sequences of Fourier coefficients of those functions defined on $(-1,1)^d$ which vanish on Ω (Proposition SM1). Now consider a smooth function g with this property. Then its Fourier expansion on $(-1,1)^d$ converges superalgebraically fast to g on $(-1,1)^d$, and therefore to zero on the domain Ω. The ratio of the norm of the truncated Fourier series on Ω divided by its norm on $(-1,1)^d$ is an upper bound for A_N, implying ill-conditioning at a superalgebraic rate.

Augmented Fourier basis. Consider the frame of Example 2. In this case, $\kappa(\boldsymbol{G}_N)$ grows algebraically fast at a rate depending on K:

PROPOSITION 9. *Let $K \in \mathbb{N}$ be fixed and consider the frame* (26). *Then $\kappa(\boldsymbol{G}_N) \gtrsim N^{2K-1}$ as $N \to \infty$.*

The intuition behind this result is as follows. The kernel of the Gram operator $\mathcal{G}$ has dimension K, and consists of sequences which are comprised of the coefficients $\{\langle p, \psi_k \rangle\}_{k=1}^{K}$ and $\{\langle -p, \varphi_n \rangle\}_{n\in\mathbb{Z}}$, where p is an arbitrary polynomial in $\mathbb{P}_K^0$. See Proposition SM4. It is possible to construct a polynomial $p \in \mathbb{P}_K^0$ which has K orders of periodic smoothness (of course, p is analytic, but it is not periodic in general). This function has Fourier coefficients $\langle p, \varphi_n \rangle$ which decay like $|n|^{-k-1}$ as $n \to \pm\infty$. Hence there is a function in H_N, i.e. the difference between p and its partial Fourier series, which is of magnitude $\mathcal{O}\left(N^{-K}\right)$ but which has $\mathcal{O}(1)$ coefficients in the frame Φ_N.

Polynomials plus modified polynomials. For the frame of Example 3 we have:

PROPOSITION 10. *Let Φ be the frame* (29) *with $w(t)$ given by* (28). *Then $\kappa(\boldsymbol{G}_N) \gtrsim 4^N$ as $N \to \infty$ up to an algebraic factor in N.*

The idea behind this result is similar to that of Example 2 (see Proposition SM7 for a description of $\mathrm{Ker}(\mathcal{G})$ in this case). We choose a polynomial $q(t) = (1+t)^{N/2-1}$ such that $w(t)q(t)$ has several orders of smoothness at $t = 0$ in spite of the algebraic singularity there. Hence, it can be well approximated by a single polynomial $p(t)$. The difference $p - wq \in \mathrm{H}_N$ is close to zero but it has $\mathcal{O}(1)$ coefficients in the frame.

Numerical illustrations of these three estimates are shown in Fig. 3. Unlike in Examples 1 and 2, the lower bound of 4^N in Example 3 does not give a good estimate of the true growth of $\kappa(\boldsymbol{G}_N)$.

5. Computing best approximations. We now turn our attention to the computation of the orthogonal projection $\mathcal{P}_N f$. Our first assertion, developed in the next subsection, is that it is impossible in general to compute $\mathcal{P}_N f$, since the coefficients of this approximation can grow arbitrarily fast with N.

5.1. Impossibility of computing best approximations. Computing $\mathcal{P}_N f = \sum_{n\in I_N} x_n \phi_n$ requires solving the ill-conditioned linear system (23). If $\boldsymbol{x} = \{x_n\}_{n\in I_N}$ and $\boldsymbol{y} = \{\langle f, \phi_n \rangle\}_{n\in I_N}$ then by Lemma 4, the Cauchy–Schwarz inequality and the
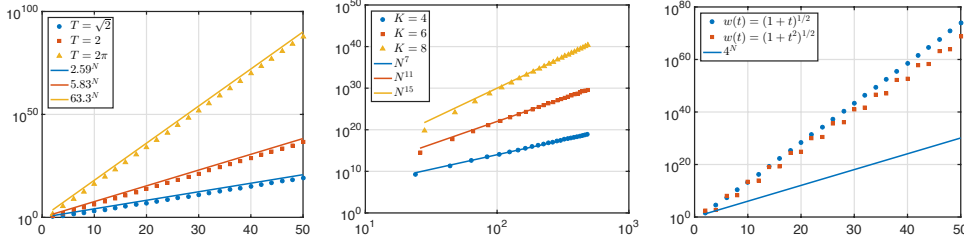
FIG. 3. *The condition numbers $\kappa(\boldsymbol{G}_N)$ for Examples 1–3 (left to right) with parameters $T = \sqrt{2}, 2, 2\pi$, $K = 4, 6, 8$ and $w(t) = \sqrt{1+t}$ respectively. The solid lines show the bounds in Propositions 8–10. Computations were carried out using additional precision.*

frame condition (13) we have

$$\|\boldsymbol{x}\| = \|\boldsymbol{G}_N^{-1}\boldsymbol{y}\| \leq \|\boldsymbol{G}_N^{-1}\|\|\boldsymbol{y}\| \leq A_N^{-1}\sqrt{\sum_{n \in I}|\langle f, \phi_n\rangle|^2} \leq A_N^{-1}\sqrt{B}\|f\|.$$

Hence, the coefficients $\boldsymbol{x}$ of the orthogonal projection may, in the worst case, grow as rapidly as A_N^{-1}. Of course, this is only an upper bound, and therefore may not be achieved for a fixed $f \in \mathrm{H}$. However, it is easy to create an example where the growth of $\|\boldsymbol{x}\|$ mirrors that of A_N^{-1}.

PROPOSITION 11. *Let $\{\varphi_n\}_{n \in \mathbb{N}}$, $g \in \mathrm{H}$, Φ and Φ_N be as in Proposition 6 and suppose that $f \in \mathrm{H}$, $\|f\| = 1$ is given by*

$$f = \frac{\sqrt{6}}{\pi}\sum_{n \in \mathbb{N}}\frac{\operatorname{sign}(\langle g, \varphi_n\rangle)}{n}\phi_n,$$

where, for $\omega \in \mathbb{C}$, $\operatorname{sign}(\omega) = \omega/|\omega|$ if $\omega \neq 0$ and $\operatorname{sign}(\omega) = 0$ otherwise. Suppose also that $\sup_{n \in \mathbb{N}}|\langle g, \varphi_n\rangle|n^2 < \infty$. If $\boldsymbol{x} = \{x_n\}_{n \in I_N}$ is the solution of (23) then

$$\|\boldsymbol{x}\| \geq \left(\frac{\pi}{\sqrt{6}}\max_{n \geq N}\{n|\langle g, \varphi_n\rangle|\}\right)^{-1}.$$

See §SM1 for a proof. Suppose now we augment a basis $\{\varphi_n\}$ with an element $g \in \mathrm{H}$ that is well approximated in the basis; for example, $|\langle g, \varphi_n\rangle| = \mathcal{O}\left(n^{-\alpha-1}\right)$ for some $\alpha \geq 1$. Then $A_N^{-1} \gtrsim N^{\alpha+1/2}$ by Proposition 6 and, if $\boldsymbol{x}$ is as in Proposition 11, then $\|\boldsymbol{x}\| \gtrsim N^{\alpha}$. Hence, there exists a fixed $f \in \mathrm{H}$ whose projection coefficients coefficients $\boldsymbol{x}$ grow rapidly as the Gram matrix $\boldsymbol{G}_N$ becomes more ill-conditioned.

Although this example is synthetic, it illustrates the general principle that the coefficients of the orthogonal projection $\mathcal{P}_N f$ in a truncated frame approximation can grow at a similar rate to that of the condition number. Hence it is generally impossible to compute $\mathcal{P}_N f$ accurately in floating-point arithmetic. To see this in a more practical setting, in Table 1 we display the coefficients for Example 1 when applied to several different functions. As is evident, only for the entire function $f(x) = \exp(x)$ is the growth of the coefficients avoided. For the other two functions, which are less smooth, we witness geometric growth of the coefficients, mirroring that of the condition number (see Proposition 8).

5.2. Truncated SVD projections. With this in mind, we now turn our attention to computing approximate projections. To regularize (23), we resort to a fa-

N	10	20	40	80	160
$f(t) = \exp(t)$	1.77e0	1.81e0	1.84e0	1.86e0	1.87e0
$f(t) = \frac{1}{1+16t^2}$	2.27e0	5.05e1	3.64e4	2.32e10	1.13e22
$f(t) = |t|^5$	2.12e-1	3.67e-1	1.76e4	7.62e26	6.09e91
$\kappa(\boldsymbol{G}_N)$	1.84e6	5.64e13	8.01e28	2.35e59	2.90e120

TABLE 1

The ℓ^2-norm $\|\boldsymbol{x}\|$ of the coefficients of the orthogonal projection $\mathcal{P}_N f$ for Example 1.

miliar approach for treating ill-conditioned and ill-posed problems (see, for example, [43, 50, 71]): compute the SVD of $\boldsymbol{G}_N$, discard all singular values below a tolerance ϵ and then find the solution $\boldsymbol{x}^\epsilon$ of the resulting system[5]. The entries of $\boldsymbol{x}^\epsilon$ are no longer the coefficients of $\mathcal{P}_N f$ but rather the projection onto a smaller space H_N^ϵ depending on ϵ. However, we shall show the the following. First, $\|\boldsymbol{x}^\epsilon\|$ is bounded (by virtue of the truncated SVD), unlike the coefficients of the orthogonal projection $\mathcal{P}_N f$. Second, despite many of the singular values having been discarded, the corresponding projection can still approximate f to high accuracy.

We first require some notation. Since $\boldsymbol{G}_N$ is positive definite its singular values $\sigma_1, \ldots, \sigma_N$ are its eigenvalues and its SVD takes the form

$$\boldsymbol{G}_N = \boldsymbol{V} \boldsymbol{\Sigma} \boldsymbol{V}^*,$$

where $\boldsymbol{V} \in \mathbb{C}^{N \times N}$ is unitary and $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_N)$ is diagonal. Write $\{\boldsymbol{v}_n : n \in I_N\}$ for the columns of $\boldsymbol{V}$, which are the left/right singular vectors of $\boldsymbol{G}_N$, i.e. $\boldsymbol{G}_N \boldsymbol{v}_n = \sigma_n \boldsymbol{v}_n$, $n \in I_N$. To each singular vector we associate a element ξ_n of H_N:

$$\xi_n = \sum_{m \in I_N} (\boldsymbol{v}_n)_m \phi_m = \mathcal{T}_N \boldsymbol{v}_n \in \mathrm{H}_N.$$

It follows from the orthogonality of $\boldsymbol{v}_n$ that the functions ξ_n are orthogonal in H:

(33) $\langle \xi_n, \xi_m \rangle = \langle \mathcal{T}_N \boldsymbol{v}_n, \mathcal{T}_N \boldsymbol{v}_m \rangle = \langle \boldsymbol{v}_n, \mathcal{T}_N^* \mathcal{T}_N \boldsymbol{v}_m \rangle = \sigma_m \langle \boldsymbol{v}_n, \boldsymbol{v}_m \rangle = \sigma_m \delta_{n,m}.$

As a result they form an orthogonal basis for H_N.

Given a tolerance $\epsilon > 0$, let $\boldsymbol{\Sigma}^\epsilon$ be the diagonal matrix with n^{th} entry σ_n if $\sigma_n > \epsilon$ and zero otherwise, and define

(34) $\boldsymbol{G}_N^\epsilon = \boldsymbol{V} \boldsymbol{\Sigma}^\epsilon \boldsymbol{V}^*.$

Then the truncated SVD coefficients $\boldsymbol{x}^\epsilon$ are $\boldsymbol{x}^\epsilon = (\boldsymbol{G}_N^\epsilon)^\dagger \boldsymbol{y} = \boldsymbol{V}(\boldsymbol{\Sigma}^\epsilon)^\dagger \boldsymbol{V}^* \boldsymbol{y}$, where $\dagger$ denotes the pseudoinverse. Note that $(\boldsymbol{\Sigma}^\epsilon)^\dagger$ is diagonal with n^{th} entry equal to $1/\sigma_n$ if $\sigma_n > \epsilon$ and 0 otherwise. We may also write $\boldsymbol{x}$ and $\boldsymbol{x}^\epsilon$ as follows:

(35) $$\boldsymbol{x} = \sum_{n \in I_N} \frac{\langle \boldsymbol{y}, \boldsymbol{v}_n \rangle}{\sigma_n} \boldsymbol{v}_n, \qquad \boldsymbol{x}^\epsilon = \sum_{\sigma_n > \epsilon} \frac{\langle \boldsymbol{y}, \boldsymbol{v}_n \rangle}{\sigma_n} \boldsymbol{v}_n,$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product on $\mathbb{C}^N$. Given $\boldsymbol{x}^\epsilon$, much like with the coefficients $\boldsymbol{x}$, we define the approximation

$$\mathcal{P}_n^\epsilon f = \mathcal{T}_N \boldsymbol{x}^\epsilon = \sum_{n \in I_N} (\boldsymbol{x}^\epsilon)_n \phi_n.$$

---

[5] We note in passing that Tikhonov regularization or similar could be used instead, with a number of changes to the results that follow.

Observe that $\langle \boldsymbol{y}, \boldsymbol{v}_n \rangle = \sum_{m \in I_N} \langle f, \phi_m \rangle \overline{(\boldsymbol{v}_n)_m} = \langle f, \xi_n \rangle$, and therefore

$$(36) \qquad \mathcal{P}_N f = \sum_{n \in I_N} \frac{\langle f, \xi_n \rangle}{\sigma_n} \xi_n, \qquad \mathcal{P}_N^\epsilon f = \sum_{\sigma_n > \epsilon} \frac{\langle f, \xi_n \rangle}{\sigma_n} \xi_n.$$

Thus, $\mathcal{P}_N^\epsilon$ is the orthogonal projection from H to H_N^ϵ, where $H_N^\epsilon = \mathrm{span}\{\xi_n : \sigma_n > \epsilon\}$.

REMARK 12. *Note that $\|\xi_n\| = \sqrt{\sigma_n}$ due to (33). Hence a singular vector $\boldsymbol{v}_n$ with small singular value σ_n corresponds to a function ξ_n with small norm in H. In other words, these functions do not contribute much to the approximation. It is therefore not all that surprising that they can be discarded without overly sacrificing accuracy.*

In Example 1, the singular vectors $\boldsymbol{v}_n$ and the functions ξ_n correspond precisely to the so-called prolate spheroidal wave sequences *and* prolate spheroidal wave functions, *introduced by Slepian, Landau and Pollak in the study of bandlimited extrapolation [61, 76]. These are a central object of study in the subfield of harmonic analysis and signal processing that focuses on time-frequency localization of signals [37, 53]. In our setting, the prolate functions corresponding to small singular values are small on Ω but large on the extended region $(-1, 1)^d \backslash \Omega$, i.e. they are approximately supported away from Ω and hence they do not influence the approximation substantially [69].*

5.3. Analysis of the truncated SVD projection $\mathcal{P}_N^\epsilon$. We now consider the error of the projection $\mathcal{P}_N^\epsilon$. The following is our main result:

THEOREM 13. *The truncated SVD projection $\mathcal{P}_N^\epsilon$ satisfies*

$$(37) \qquad \|f - \mathcal{P}_N^\epsilon f\| \leq \inf \left\{ \|f - \mathcal{T}_N \boldsymbol{z}\| + \sqrt{\epsilon} \|\boldsymbol{z}\| : \boldsymbol{z} \in \mathbb{C}^N \right\}.$$

Proof. Let $\boldsymbol{z} \in \mathbb{C}^N$. Since $\mathcal{P}_N^\epsilon$ is the orthogonal projection onto H_N^ϵ, we have

$$\|f - \mathcal{P}_N^\epsilon f\| \leq \|f - \mathcal{P}_N^\epsilon \mathcal{T}_N \boldsymbol{z}\| \leq \|f - \mathcal{T}_N \boldsymbol{z}\| + \|\mathcal{T}_N \boldsymbol{z} - \mathcal{P}_N^\epsilon \mathcal{T}_N \boldsymbol{z}\|.$$

Note that $\mathcal{T}_N \boldsymbol{z} = \mathcal{P}_N \mathcal{T}_N \boldsymbol{z}$ since $\mathcal{T}_N \boldsymbol{z} \in H_N$. Hence (36) and the orthogonality of the ξ_n's gives

$$\|\mathcal{T}_N \boldsymbol{z} - \mathcal{P}_N^\epsilon \mathcal{T}_N \boldsymbol{z}\|^2 = \left\| \sum_{\sigma_n \leq \epsilon} \frac{\langle \mathcal{T}_N \boldsymbol{z}, \xi_n \rangle}{\sigma_n} \xi_n \right\|^2 = \sum_{\sigma_n \leq \epsilon} \frac{|\langle \mathcal{T}_N \boldsymbol{z}, \xi_n \rangle|^2}{\sigma_n}.$$

Observe that $\langle \mathcal{T}_N \boldsymbol{z}, \xi_n \rangle = \langle \mathcal{T}_N \boldsymbol{z}, \mathcal{T}_N \boldsymbol{v}_n \rangle = \langle \boldsymbol{z}, \boldsymbol{G}_N \boldsymbol{v}_n \rangle = \sigma_n \langle \boldsymbol{z}, \boldsymbol{v}_n \rangle$ and therefore

$$\|\mathcal{T}_N \boldsymbol{z} - \mathcal{P}_N^\epsilon \mathcal{T}_N \boldsymbol{z}\|^2 = \sum_{\sigma_n \leq \epsilon} \sigma_n |\langle \boldsymbol{z}, \boldsymbol{v}_n \rangle|^2 \leq \epsilon \sum_{n \in I_N} |\langle \boldsymbol{z}, \boldsymbol{v}_n \rangle|^2 = \epsilon \|\boldsymbol{z}\|^2,$$

where in the last step we use the fact that the vectors $\{\boldsymbol{v}_n\}_{n \in I_N}$ are orthonormal. □

This theorem establishes the claim made earlier in the paper: the convergence of the projection $\mathcal{P}_N^\epsilon f$ is dictated by how well f can be approximated by coefficients $\boldsymbol{z}$ with small norm. Note that this situation is markedly different to the case of the projection $\mathcal{P}_N f$, wherein the analogous error bound is simply $\|f - \mathcal{P}_N f\| \leq \inf\{\|f - \mathcal{T}_N \boldsymbol{z}\| : \boldsymbol{z} \in \mathbb{C}^N\}$. As we discuss in §5.4, the appearance of the term $\sqrt{\epsilon}\|\boldsymbol{z}\|$ can change the behaviour of the error in a key way. We note in passing however that if $\boldsymbol{a} = \{\langle f, \mathcal{S}^{-1} \phi_n \rangle\}_{n \in I}$ are the frame coefficients of f, then setting $\boldsymbol{z} = \{a_n\}_{n \in I_N}$ in (37) and allowing $N \to \infty$ gives

$$(38) \qquad \limsup_{N \to \infty} \|f - \mathcal{P}_N^\epsilon f\| \leq \sqrt{\epsilon} \|\boldsymbol{a}\| \leq \sqrt{\epsilon/A} \|f\|,$$

where the second inequality is due to (17). Hence the projection $\mathcal{P}_N^\epsilon f$ eventually approximates f to within a factor of $\sqrt{\epsilon}$.

We next consider the behaviour of the coefficients $\boldsymbol{x}^\epsilon$:

THEOREM 14. *The coefficients $\boldsymbol{x}^\epsilon$ of the truncated SVD projection $\mathcal{P}_N^\epsilon$ satisfy*

$$(39) \qquad \|\boldsymbol{x}^\epsilon\| \leq \inf\left\{1/\sqrt{\epsilon}\|f - \mathcal{T}_N \boldsymbol{z}\| + \|\boldsymbol{z}\| : \boldsymbol{z} \in \mathbb{C}^N\right\}.$$

Moreover, if $\boldsymbol{a}_N^\epsilon \in \ell^2(I)$ is the extension of $\boldsymbol{x}^\epsilon$ by zero, then

$$(40) \qquad \|\boldsymbol{a} - \boldsymbol{a}_N^\epsilon\| \leq \left(1 + \sqrt{B/\epsilon}\right)\sqrt{\sum_{n\in I\setminus I_N} |a_n|^2} + \sqrt{\epsilon/A}\|\boldsymbol{a}\|,$$

where $\boldsymbol{a} = \{\langle f, \mathcal{S}^{-1}\phi_n\rangle\}_{n\in I}$ are the frame coefficients of f.

Proof. For the first part, we use (35) to write

$$\boldsymbol{x}^\epsilon = \sum_{\sigma_n > \epsilon} \frac{\langle f, \xi_n\rangle}{\sigma_n}\boldsymbol{v}_n = \sum_{\sigma_n > \epsilon} \frac{\langle f - \mathcal{T}_N \boldsymbol{z}, \xi_n\rangle}{\sigma_n}\boldsymbol{v}_n + \sum_{\sigma_n > \epsilon} \frac{\langle \mathcal{T}_N \boldsymbol{z}, \xi_n\rangle}{\sigma_n}\boldsymbol{v}_n.$$

Consider the first term on the right-hand side. By (33) and (36) we have

$$\left\|\sum_{\sigma_n > \epsilon} \frac{\langle f - \mathcal{T}_N \boldsymbol{z}, \xi_n\rangle}{\sigma_n}\boldsymbol{v}_n\right\|^2 = \sum_{\sigma_n > \epsilon} \frac{|\langle f - \mathcal{T}_N \boldsymbol{z}, \xi_n\rangle|^2}{\sigma_n^2} \leq \frac{1}{\epsilon}\|\mathcal{P}_N^\epsilon (f - \mathcal{T}_N \boldsymbol{z})\|^2 \leq \frac{1}{\epsilon}\|f - \mathcal{T}_N \boldsymbol{z}\|^2.$$

For the second term, we notice that $\langle \mathcal{T}_N \boldsymbol{z}, \xi_n\rangle = \sigma_n \langle \boldsymbol{z}, \boldsymbol{v}_n\rangle$, and therefore

$$\left\|\sum_{\sigma_n > \epsilon} \frac{\langle \mathcal{T}_N \boldsymbol{z}, \xi_n\rangle}{\sigma_n}\boldsymbol{v}_n\right\|^2 = \sum_{\sigma_n > \epsilon} |\langle \boldsymbol{z}, \boldsymbol{v}_n\rangle|^2 \leq \|\boldsymbol{z}\|^2.$$

Combining these two bounds now gives the first result. For the second result, we first let $\boldsymbol{a}_N \in \mathbb{C}^N$ be the vector with n^{th} entry $a_n = \langle f, \mathcal{S}^{-1}\phi_n\rangle$ for $n \in I_N$. Then

$$\|\boldsymbol{a} - \boldsymbol{a}_N^\epsilon\| \leq \sqrt{\sum_{n\in I\setminus I_N} |a_n|^2} + \|\boldsymbol{a}_N - \boldsymbol{x}^\epsilon\|.$$

Hence it suffices to estimate $\|\boldsymbol{a}_N - \boldsymbol{x}^\epsilon\|$. For this, we note that $f = \mathcal{S}\mathcal{S}^{-1}f = \mathcal{S}_N\mathcal{S}^{-1}f + (\mathcal{S} - \mathcal{S}_N)\mathcal{S}^{-1}f$. Since $\mathcal{S}_N$ is self-adjoint and $\mathcal{S}_N\xi_n = \mathcal{T}_N \mathcal{T}_N^* \mathcal{T}_N \boldsymbol{v}_n = \sigma_n \mathcal{T}_N \boldsymbol{v}_n = \sigma_n \xi_n$ we have

$$\langle f, \xi_n\rangle = \langle \mathcal{S}_N\mathcal{S}^{-1}f, \xi_n\rangle + \langle (\mathcal{S} - \mathcal{S}_N)\mathcal{S}^{-1}f, \xi_n\rangle = \sigma_n\langle \mathcal{S}^{-1}f, \xi_n\rangle + \langle (\mathcal{S} - \mathcal{S}_N)\mathcal{S}^{-1}f, \xi_n\rangle.$$

Therefore

$$(41) \qquad \boldsymbol{x}^\epsilon = \sum_{\sigma_n > \epsilon} \frac{\langle f, \xi_n\rangle}{\sigma_n}\boldsymbol{v}_n = \sum_{\sigma_n > \epsilon} \langle \mathcal{S}^{-1}f, \xi_n\rangle\boldsymbol{v}_n + \sum_{\sigma_n > \epsilon} \frac{1}{\sigma_n}\langle (\mathcal{S} - \mathcal{S}_N)\mathcal{S}^{-1}f, \xi_n\rangle\boldsymbol{v}_n.$$

Conversely, since $\boldsymbol{a}_N = \mathcal{T}_N^*\mathcal{S}^{-1}f$ we have $\langle \boldsymbol{a}_N, \boldsymbol{v}_n\rangle = \langle \mathcal{S}^{-1}f, \mathcal{T}_N \boldsymbol{v}_n\rangle = \langle \mathcal{S}^{-1}f, \xi_n\rangle$. Hence

$$(42) \qquad \boldsymbol{a}_N = \sum_{n\in I_N} \langle \boldsymbol{a}_N, \boldsymbol{v}_n\rangle\boldsymbol{v}_n = \sum_{n\in I_N} \langle \mathcal{S}^{-1}f, \xi_n\rangle\boldsymbol{v}_n.$$

Combining (41) and (42) now gives

$$
(43) \qquad \|\boldsymbol{a}_N - \boldsymbol{x}^\epsilon\| \le \left\| \sum_{\sigma_n \le \epsilon} \langle \mathcal{S}^{-1} f, \xi_n \rangle \boldsymbol{v}_n \right\| + \left\| \sum_{\sigma_n > \epsilon} \frac{1}{\sigma_n} \langle (\mathcal{S} - \mathcal{S}_N) \mathcal{S}^{-1} f, \xi_n \rangle \boldsymbol{v}_n \right\|.
$$

Consider the first term. By orthogonality

$$
(44) \qquad \left\| \sum_{\sigma_n \le \epsilon} \langle \mathcal{S}^{-1} f, \xi_n \rangle \boldsymbol{v}_n \right\|^2 \le \epsilon \sum_{\sigma_n \le \epsilon} \frac{1}{\sigma_n} |\langle \mathcal{S}^{-1} f, \xi_n \rangle|^2 \le \epsilon \|\mathcal{S}^{-1} f\|^2 \le \epsilon/A \|\boldsymbol{a}\|^2.
$$

Here the final inequality follows from (13). Now consider the second term:

$$
\begin{aligned}
\left\| \sum_{\sigma_n > \epsilon} \frac{1}{\sigma_n} \langle (\mathcal{S} - \mathcal{S}_N) \mathcal{S}^{-1} f, \xi_n \rangle \boldsymbol{v}_n \right\|^2 &= \sum_{\sigma_n > \epsilon} \frac{1}{\sigma_n^2} |\langle (\mathcal{S} - \mathcal{S}_N) \mathcal{S}^{-1} f, \xi_n \rangle|^2 \\
&\le \frac{1}{\epsilon} \sum_{\sigma_n > \epsilon} \frac{1}{\sigma_n} |\langle (\mathcal{S} - \mathcal{S}_N) \mathcal{S}^{-1} f, \xi_n \rangle|^2 \\
&\le \frac{1}{\epsilon} \|(\mathcal{S} - \mathcal{S}_N) \mathcal{S}^{-1} f\|^2.
\end{aligned}
$$

Observe that

$$
\|(\mathcal{S} - \mathcal{S}_N) \mathcal{S}^{-1} f\| = \left\| \sum_{n \in I \setminus I_N} a_n \phi_n \right\| = \sup_{\substack{g \in \mathrm{H} \\ g \ne 0}} \left\{ \frac{\left| \sum_{n > N} a_n \overline{\langle g, \phi_n \rangle} \right|}{\|g\|} \right\} \le \sqrt{B} \sqrt{\sum_{n \in I \setminus I_N} |a_n|^2},
$$

and therefore

$$
\left\| \sum_{\sigma_n > \epsilon} \frac{1}{\sigma_n} \langle (\mathcal{S} - \mathcal{S}_N) \mathcal{S}^{-1} f, \xi_n \rangle \boldsymbol{v}_n \right\|^2 \le B/\epsilon \sum_{n \in I \setminus I_N} |a_n|^2.
$$

Substituting this and (44) into (43) gives the result. □

Theorems 13 and 14 show a rather surprising conclusion. Despite severe ill-conditioning of the Gram matrix, which led us to discard all of its singular values of size less than ϵ, one still gets convergence of $\mathcal{P}_N^\epsilon f$ to within $\sqrt{\epsilon}$ of f. Moreover, although the coefficients $\boldsymbol{x}^\epsilon$ may initially grow large (due to the $1/\sqrt{\epsilon}$ factor in (39)), they too eventually converge to within $\sqrt{\epsilon}$ of the frame coefficients of f.

The underlying reason for this, as detailed in the following theorem, is that the absolute condition number of the mapping $\boldsymbol{y} = \{\langle f, \phi_n \rangle\}_{n \in I_N} \mapsto \mathcal{P}_N^\epsilon f$ from the data $\boldsymbol{y}$ to the projection $\mathcal{P}_N^\epsilon f$ is much smaller – precisely, on the order of $1/\sqrt{\epsilon}$ as opposed to $1/\epsilon$ – than the absolute condition number of the mapping from $\boldsymbol{y}$ to the coefficients $\boldsymbol{x}^\epsilon$ (which is just the minimal singular value of the SVD truncated Gram matrix $\boldsymbol{G}_N^\epsilon$; see (34)). Hence, whilst errors in the coefficients may be on the order of $1/\epsilon$, they result in much smaller errors in the projection $\mathcal{P}_N^\epsilon$.

THEOREM 15. *The absolute condition number of the mapping $\mathbb{C}^N \to \mathrm{H}_N^\epsilon$, $\boldsymbol{y} \mapsto \mathcal{T}_N(\boldsymbol{G}_N^\epsilon)^\dagger \boldsymbol{y}$ satisfies $\kappa \le \min\left\{ 1/\sqrt{\sigma_{\min}(\boldsymbol{G}_N)}, 1/\sqrt{\epsilon} \right\}$, where $\sigma_{\min}(\boldsymbol{G}_N)$ is the minimal singular value of the Gram matrix $\boldsymbol{G}_N$.*

Proof. By linearity, the condition number of the mapping $\boldsymbol{y} \mapsto \mathcal{T}_N(\boldsymbol{G}_N^\epsilon)^\dagger \boldsymbol{y}$ is $\kappa = \max_{\substack{\boldsymbol{y} \in \mathbb{C}^N \\ \|\boldsymbol{y}\|=1}} \|\mathcal{T}_N(\boldsymbol{G}_N^\epsilon)^\dagger \boldsymbol{y}\|$. We have

$$\|\mathcal{T}_N(\boldsymbol{G}_N^\epsilon)^\dagger \boldsymbol{y}\|^2 = \langle \boldsymbol{y}, (\boldsymbol{G}_N^\epsilon)^\dagger \boldsymbol{G}_N (\boldsymbol{G}_N^\epsilon)^\dagger \boldsymbol{y}\rangle = \langle \boldsymbol{y}, (\boldsymbol{G}_N^\epsilon)^\dagger \boldsymbol{y}\rangle = \sum_{\sigma_n > \epsilon} \frac{|\langle \boldsymbol{y}, \boldsymbol{v}_n\rangle|^2}{\sigma_n}.$$

This gives $\|\mathcal{T}_N(\boldsymbol{G}_N^\epsilon)^\dagger \boldsymbol{y}\|^2 \leq 1/\min\{\sigma_n : \sigma_n > \epsilon\}\|\boldsymbol{y}\|^2$, and the result follows. □

REMARK 16. *Theorems 13 and 14 assert convergence to within $\sqrt{\epsilon}$ of f only. Using spectral theory techniques it can be shown that $\boldsymbol{a}_N^\epsilon \to \boldsymbol{a}$ as $N \to \infty$ [51, Thm. 5.17], i.e. the regularized coefficients converge to the frame coefficients in the canonical dual frame. Since $\|f - \mathcal{P}_N^\epsilon f\| = \|\mathcal{T}(\boldsymbol{a} - \boldsymbol{a}_N^\epsilon)\| \leq \sqrt{B}\|\boldsymbol{a} - \boldsymbol{a}_N^\epsilon\|$ this also gives $\mathcal{P}_N^\epsilon f \to f$, albeit at a rate that can be arbitrarily slow. Of course, in finite precision calculations convergence beyond $\mathcal{O}(\sqrt{\epsilon})$ will not be expected, due to the condition number of the mapping (Theorem 15). We note also that convergence down to $\mathcal{O}(\sqrt{\epsilon})$ has previously been observed empirically in [51, 81]. The main contribution of Theorems 13 and 14 is that they provide explicit bounds which can be used to estimate the rate of decay of these errors in the regime where they are larger than $\sqrt{\epsilon}$.*

5.4. Discussion and examples. We now consider Theorems 13–15 in relation to the examples of §3. We focus on two issues: (i) the convergence of the projection $\mathcal{P}_N^\epsilon f$, and (ii) the behaviour of the coefficients $\boldsymbol{x}^\epsilon$. For proofs of the various results presented, we refer to the supplementary material.

First, notice that for small N – specifically, for N such that $\sigma_{\min}(\boldsymbol{G}_N) \geq \epsilon$ – we have $\mathcal{P}_N^\epsilon = \mathcal{P}_N$. Hence, the truncated SVD projection initially behaves like the exact projection $\mathcal{P}_N$. However, beyond this point the convergence begins to differ. If $\boldsymbol{x} \in \mathbb{C}^N$ are the coefficients of $\mathcal{P}_N f$ then setting $\boldsymbol{z} = \boldsymbol{x}$ in (37) gives

$$\|f - \mathcal{P}_N^\epsilon f\| \leq \|f - \mathcal{P}_N f\| + \sqrt{\epsilon}\|\boldsymbol{x}\|.$$

As discussed in §5.1, the term $\|\boldsymbol{x}\|$ often grows rapidly in N. Hence as N increases, the right-hand side of the above inequality may begin to diverge.

However, since Φ is a frame there are infinitely-many sequences of coefficients $\boldsymbol{c} \in \ell^2(I)$ such that $f = \mathcal{T}\boldsymbol{c}$. Suppose that $\boldsymbol{z} = \{c_n\}_{n \in I_N}$. Then Theorem 13 gives

$$(45) \qquad \|f - \mathcal{P}_N^\epsilon f\| \leq \|f - \mathcal{T}_N \boldsymbol{c}\| + \sqrt{\epsilon}\|\boldsymbol{c}\| = \left\|\sum_{n \in I \backslash I_N} c_n \phi_n\right\| + \sqrt{\epsilon}\|\boldsymbol{c}\|.$$

The term $\sqrt{\epsilon}\|\boldsymbol{c}\|$ is independent of N, while the other term tends to zero as $N \to \infty$. Hence, the rate of decay of the error down to $\sqrt{\epsilon}$ is bounded by how well f can be expanded in the frame using coefficients $\boldsymbol{c}$ with small norms. In the examples below, we show that there always exist coefficient sequences $\boldsymbol{c}$ that achieve favourable rates of decay. Therefore, while the truncated SVD projection $\mathcal{P}_N^\epsilon f$ may not achieve the same error decay as the exact projection $\mathcal{P}_N f$, we can often expect good accuracy.

Fourier frames for complex geometries. Consider the frame of Example 1. We have the following:

PROPOSITION 17. *Let $\Omega \subseteq (-1, 1)^d$ be a compact domain with Lipschitz boundary and consider the frame (24). If $f \in \mathrm{H}^{kd}(\Omega)$ then there exists a $\boldsymbol{c} \in \ell^2(I)$ such that*

$$\|f - \mathcal{T}_N \boldsymbol{c}\| \leq C N^{-k}\|f\|_{\mathrm{H}^{kd}(-1,1)^d}, \qquad \|\boldsymbol{c}\| \leq C\|f\|_{\mathrm{H}^{kd}(-1,1)^d},$$

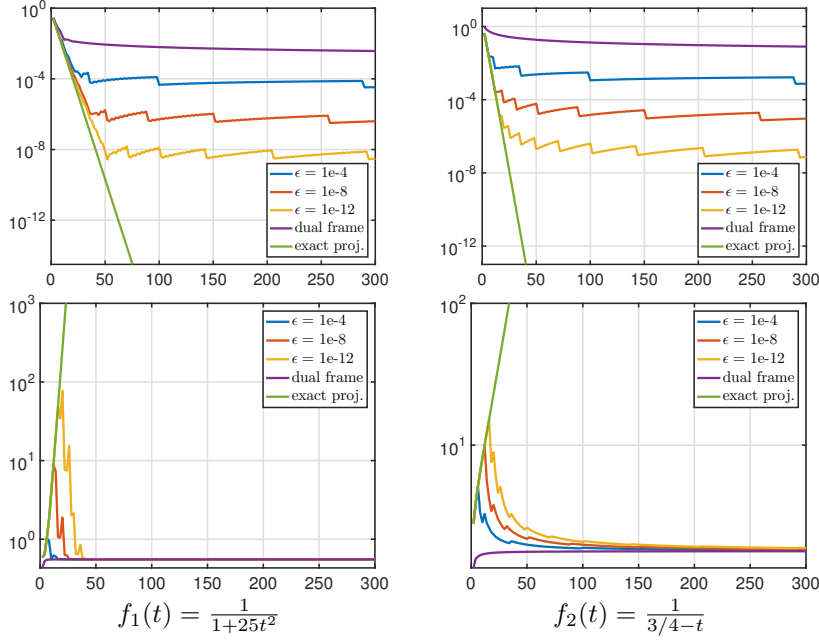$$f_1(t) = \frac{1}{1+25t^2} \qquad\qquad f_2(t) = \frac{1}{3/4-t}$$

FIG. 4. *Comparison of the exact projection $\mathcal{P}_N f$, the truncated SVD projections $\mathcal{P}_N^\epsilon f$ and the canonical dual frame expansion $\sum_{n\in I_N} a_n \phi_n$ for Example 1 with $T = 2$. Top row: L^2-norm errors versus N. Bottom row: the norms of the coefficient vectors versus N. The exact projection was computed using additional precision. All other approximations were computed in double precision.*

where $C > 0$ is independent of f and N. In particular, for the exact projection

$$\|f - \mathcal{P}_N f\| \le C N^{-k} \|f\|_{\mathrm{H}^{kd}(-1,1)^d},$$

and for the regularized projection,

$$(46) \qquad\qquad \|f - \mathcal{P}_N^\epsilon f\| \le C\left(N^{-k} + \sqrt{\epsilon}\right) \|f\|_{\mathrm{H}^k(-1,1)^d}.$$

This proposition asserts that there are bounded coefficient vectors for which the error $\|f - \mathcal{T}_N \boldsymbol{c}\|$ decays at an arbitrarily-high algebraic rate. In particular, the error of the exact projection decays spectrally fast in N. On the other hand, for the regularized projection we have the following. First, for smooth functions f, the error decays rapidly in N when $\|f - \mathcal{P}_N^\epsilon f\| \gg \sqrt{\epsilon}$. Second, if the kth derivatives of f grow rapidly in k then the rate of decrease of the error may lessen as $\|f - \mathcal{P}_N^\epsilon f\|$ approaches $\sqrt{\epsilon}$, due to the presence of the Sobolev norm in the error bound (46). This effect is shown in Fig. 4. Note that the derivatives of the function f_1 grow slowly with k, whereas the derivatives of f_2 grow much more rapidly. As predicted by (46), there is substantially less effect in replacing $\mathcal{P}_N$ by the regularized projection $\mathcal{P}_N^\epsilon$ in the case of f_1 than in the case of f_2. This aside, Fig. 4 also shows the slow convergence of the canonical dual frame expansion (16), thus confirming the discussion in §3, and the norms of the various coefficient vectors. These results are in good agreement with Theorem 14: the coefficients of $\mathcal{P}_N^\epsilon f$ initially grow large, but after a certain point they begin to decay to the limiting value $\|\boldsymbol{a}\|$.

Augmented Fourier basis. Consider the frame of Example 2. In this case we

$$f_1(t) = \frac{1}{10-9t}$$

$$f_2(t) = e^{\sin(3t+1/2)}\sqrt{1+t} + \cos(5t)$$

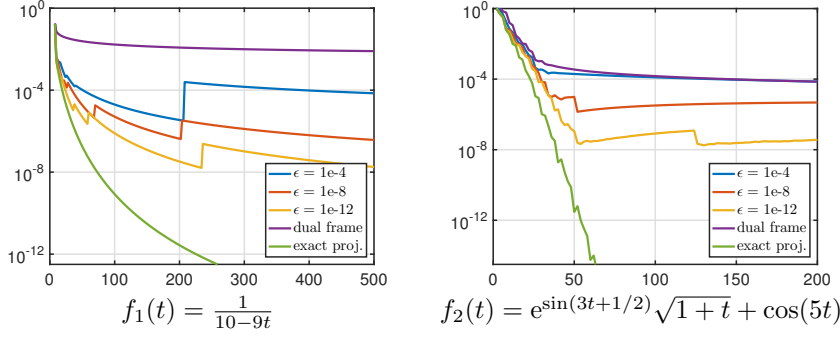FIG. 5. L^2-norm errors versus N for the exact projection $\mathcal{P}_N f$, the truncated SVD projections $\mathcal{P}_N^\epsilon f$ and the canonical dual frame expansion $\sum_{n \in I_N} a_n \phi_n$ for Example 2 with $K = 8$ (left) and Example 3 with $\alpha = 1/2$ (right). The exact projection was computed using additional precision. All other approximations were computed in double precision.

observe algebraic convergence at a rate determined by the number of polynomials K added to the Fourier basis. See also Fig. 5.

PROPOSITION 18. *Let $K \in \mathbb{N}$ be fixed and consider the frame* (26). *If $f \in \mathrm{H}^k(-1,1)$ for $0 \le k \le K$ then there exists a $\mathbf{c} \in \ell^2(I)$ such that*

$$\|f - \mathcal{T}_N \mathbf{c}\| \le C N^{-k} \|f\|_{\mathrm{H}^k(-1,1)}, \qquad \|\mathbf{c}\| \le C \|f\|_{\mathrm{H}^k(-1,1)},$$

where $C > 0$ is independent of f and N. In particular,

$$\|f - \mathcal{P}_N f\| \le C N^{-k} \|f\|_{\mathrm{H}^k(-1,1)},$$
$$\|f - \mathcal{P}_N^\epsilon f\| \le C \left(N^{-k} + \sqrt{\epsilon}\right) \|f\|_{\mathrm{H}^k(-1,1)}.$$

Polynomials plus modified polynomials. For the frame of Example 3 we have:

PROPOSITION 19. *Let $\mathcal{Q}_N : \mathrm{L}^2(-1,1) \to \mathrm{L}^2(-1,1)$ be the orthogonal projection onto $\mathbb{P}_{N/2-1}$. If $f(t) = w(t)g(t) + h(t)$, then there exists a $\mathbf{c} \in \ell^2(I)$ such that*

$$\|f - \mathcal{T}_N \mathbf{c}\| \le w_{\max} \|g - \mathcal{Q}_N g\| + \|h - \mathcal{Q}_N h\|, \qquad \|\mathbf{c}\| \le \|g\| + \|h\|,$$

where $w_{\max} = \operatorname{ess\,sup}_{t \in (-1,1)} |w(t)|$. In particular,

$$\|f - \mathcal{P}_N f\| \le w_{\max} \|g - \mathcal{Q}_N g\| + \|h - \mathcal{Q}_N h\|,$$
$$\|f - \mathcal{P}_N^\epsilon f\| \le w_{\max} \|g - \mathcal{Q}_N g\| + \|h - \mathcal{Q}_N h\| + \sqrt{\epsilon} \left(\|g\| + \|h\|\right).$$

This result implies the convergence of the regularized projection $\mathcal{P}_N^\epsilon f$ is spectral in the factors g and h. In particular, if g and h are infinitely differentiable then one has superalgebraic convergence down to $\sqrt{\epsilon}$, and if g and h are analytic, one has geometric convergence down to $\sqrt{\epsilon}$. See Fig. 5. This figure suggests geometric convergence, in agreement with Proposition 19, but with a somewhat reduced exponent over that of the exact projection $\mathcal{P}_N f$. In other words, there exist coefficient vectors in the frame which yield faster geometric convergence, but are too large to be obtained as solutions of the regularized system.

6. Well-conditioned frame approximations. As shown in the previous section, the accuracy of the numerical frame projection $\mathcal{P}_N^\epsilon$ is limited to $\mathcal{O}(\sqrt{\epsilon})$, and the relevant condition number is at most $1/\sqrt{\epsilon}$. We close this paper with a brief description of an approximation that achieves $\mathcal{O}(\epsilon)$ accuracy and has a bounded condition number independent of ϵ. This approximation is based on oversampling. Specifically, rather than solving the square system (23), we consider the $M \times N$ system

$$(47) \qquad \boldsymbol{G}_{M,N}\boldsymbol{x} \approx \boldsymbol{y}, \qquad \boldsymbol{y} = \{\langle f, \phi_n \rangle\}_{n \in I_M},$$

where $\boldsymbol{G}_{M,N} = \{\langle \phi_n, \phi_m \rangle\}_{m \in I_M, n \in I_N} \in \mathbb{C}^{M \times N}$. Note that where $\boldsymbol{G}_N$ corresponds to the finite section of the Gram operator $\mathcal{G}$, $\boldsymbol{G}_{M,N}$ corresponds to a so-called *uneven section*. Uneven sections are known to be useful alternatives to finite sections in computational spectral theory [14, 47, 48, 49, 63] and, more recently, sampling theory [3, 4, 5]. Much the same is true in this instance.

Since Gram operators of frames are singular, $\boldsymbol{G}_{M,N}$ remains ill-conditioned even when $M \geq N$. Hence we consider the regularized projection

$$\boldsymbol{x}^\epsilon = (\boldsymbol{G}_{M,N}^\epsilon)^\dagger \boldsymbol{y}, \qquad \mathcal{P}_{M,N}^\epsilon f = \sum_{n \in I_N} (\boldsymbol{x}^\epsilon)_n \phi_n.$$

where $\boldsymbol{G}_{M,N}^\epsilon$ is obtained by discarding all its singular values of $\boldsymbol{G}_{M,N}$ below ϵ.

Given a sufficient amount of oversampling, this approximation is well conditioned and achieves $\mathcal{O}(\epsilon)$ accuracy. Specifically, it can be shown that the condition number of the mapping $\boldsymbol{y} \mapsto \mathcal{T}_N(\boldsymbol{G}_{M,N}^\epsilon)^\dagger \boldsymbol{y}$ is bounded by $C_{M,N}^\epsilon/\sqrt{A}$, and approximation $\mathcal{P}_{M,N}^\epsilon f$ satisfies the error bound

$$(48) \quad \|f - \mathcal{P}_{M,N}^\epsilon f\| \leq \left(1 + \sqrt{B/A}C_{M,N}^\epsilon\right) \inf\left\{\|f - \mathcal{T}_N \boldsymbol{z}\| + \epsilon/\sqrt{B}\|\boldsymbol{z}\| : \boldsymbol{z} \in \mathbb{C}^N\right\}.$$

Here $C_{M,N}^\epsilon$ is a constant depending on Φ, ϵ, M and N. It satisfies

$$\limsup_{M \to \infty} C_{M,N}^\epsilon \leq 1,$$

for any fixed N and ϵ. In other words, this constant can be made arbitrarily close to one with sufficiently large M. Note that the inequality (48) is similar to the error bound (37) for the projection $\mathcal{P}_N^\epsilon f$, except for appearance of ϵ in place of $\sqrt{\epsilon}$. Hence the error now decays down to $\mathcal{O}(\epsilon)$, with, as before, the rate of decay of the error being dictated by existence of expansions in the frame with small-norm coefficients.

The proofs of these results are given in a companion paper [7], which develops on this paper by introducing a general framework for frame approximations from 'indirect' data. This framework also includes discrete function samples, for example, which are typically much more convenient to work with than (47) since they do not require evaluations of inner products.

In Fig. 6 we present several numerical results for the Examples 1–3 based on (47). These results illustrate that with a mild amount oversampling one can obtain a much more accurate numerical frame approximation than when $M = N$, which is the case considered previously (see [9] for some further analysis in the case of Example 1). Similar results can also be obtained via least-squares fitting with discrete function samples taken, for example, on an equally-spaced grid of $M \approx 2N$ points [7]. A fast algorithm to do so is given in [70], and an illustration is shown in Fig. 1.
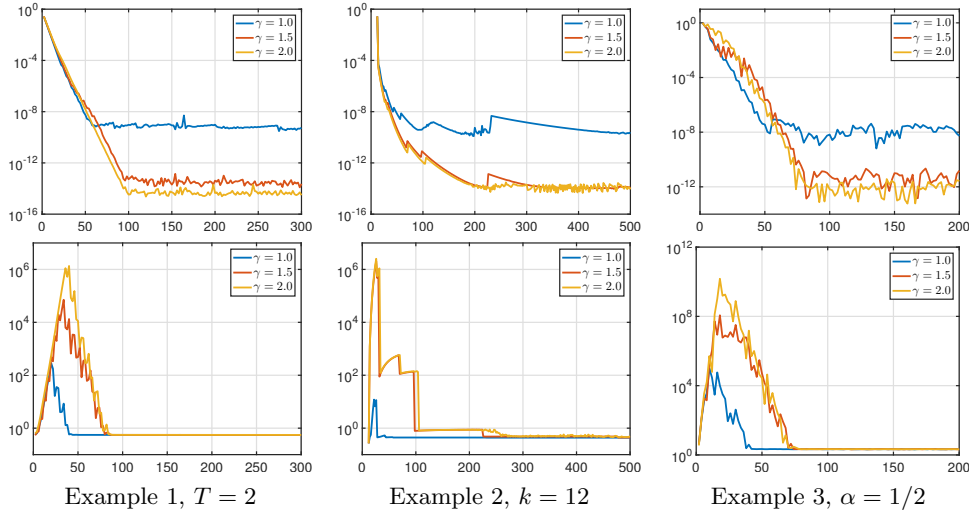
FIG. 6. *Comparison of the projection $\mathcal{P}^{\epsilon}_{\gamma N, N} f$ for different values of the parameter γ. Top row: L^2-norm errors versus N. Bottom row: the norms of the coefficient vectors versus N. The functions used were $f(t) = \frac{1}{1+25t^2}$ (left), $f(t) = \frac{1}{5-4t}$ (middle) and $f(t) = e^{\sin(3t+1/2)}\sqrt{1+t} + \cos(5t)$ (right).*

7. Conclusions and further research. The concern of this paper has been computing numerical approximations using frames, in particular, orthogonal projections in the span of the first N frame elements. There are four main conclusions. First, the Gram matrices of truncated frames are necessarily ill-conditioned. As shown, such ill-conditioning is due to approximation of the singular Gram operator by its finite section, and can be arbitrarily bad depending on the frame. Second, the orthogonal projection typically cannot be computed, since its coefficients generally grow rapidly in N. However, using regularization it is possible to find an approximation in a truncated frame whose coefficients are bounded (by at most $\mathcal{O}\left(1/\sqrt{\epsilon}\right)$ and eventually are $\mathcal{O}(1)$), has condition number at most $1/\sqrt{\epsilon}$, and which converges down to an error of $\mathcal{O}\left(\sqrt{\epsilon}\right)$. The convergence of this approximation depends on how well f can be approximated by finite expansions in the frame with small-norm coefficients. Fourth, and finally, using oversampling one can construct an approximation that has $\mathcal{O}(1)$ condition number and is convergent down to $\mathcal{O}\left(\epsilon\right)$.

The overall conclusion of this paper is that satisfactory approximations can be computed in certain finite systems with near-linear dependencies. We stress that the frame condition is important in this regard. The monomials $\Phi_N = \{1, x, x^2, \ldots x^{N-1}\}$ are nearly-linearly dependent for large N, but may not give good numerical approximations, since a smooth function does not necessarily have a representation in this system with small-norm coefficients. On the other hand, a frame guarantees at least one approximation with small-norm coefficients – namely, the truncated canonical dual frame expansion – although, as seen, better approximations often exist. At this stage we stress again that the approach developed in this paper is recommended only when the dual frame expansion gives unsatisfactory convergence, such as in Examples 1–3. Otherwise, approximation via this expansion is arguably preferable (recall §1.6).

There are a number of topics not considered in this paper. First, our main results apply to general frames. Yet there may be theoretical and practical advantages to considering classes of frames with particular properties; for instance, so-called *localized*

frames (see [11, 45, 46, 78] and references therein). Second, we have used a fixed parameter ϵ throughout. However, adaptive (i.e. function dependent) choices would likely be beneficial. This warrants further study. Third, since the focus of this paper has been on general frames, we have not considered fast computations (which is more specific to the frame employed). A fast algorithm for computing Fourier extensions in the one-dimensional setting was introduced in [66] for the special case where the extension interval has exactly twice the length of the original interval. A more recent alternative that also generalizes to higher dimensions is described in [69, 70]. Crucial elements in the latter approach include the link to the theory of bandlimited functions, the special prolate spheroidal wave functions and a phenomenon called the *plunge region* in sampling theory. It seems these elements may generalize to other types of frames, which is a topic that will be considered in future work. Fourth and finally, we have not discussed the accuracy of computing the SVD of $\boldsymbol{G}_N$ in floating-point arithmetic, and its effect on the numerical projection. Numerical experiments suggest this does not have a substantial impact on the error, but since $\boldsymbol{G}_N$ is severely ill-conditioned a careful analysis should be carried out. We expect the structure of the singular values (in particular, their tendency to divide into 'good' singular value away from zero and 'bad' singular values near zero) is important in this regard.

REFERENCES

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

[2] B. Adcock, *Convergence acceleration of modified Fourier series in one or more dimensions*, Math. Comp., 80 (2011), pp. 225–261.

[3] B. Adcock and A. C. Hansen, *A generalized sampling theorem for stable reconstructions in arbitrary bases*, J. Fourier Anal. Appl., 18 (2012), pp. 685–716.

[4] B. Adcock and A. C. Hansen, *Generalized sampling and infinite-dimensional compressed sensing*, Found. Comput. Math., 16 (2016), pp. 1263–1323.

[5] B. Adcock, A. C. Hansen, and C. Poon, *Beyond consistent reconstructions: optimality and sharp bounds for generalized sampling, and application to the uniform resampling problem.*, SIAM J. Math. Anal., 45 (2013), pp. 3114–3131.

[6] B. Adcock and D. Huybrechs, *On the resolution power of Fourier extensions for oscillatory functions*, J. Comput. Appl. Math., 260 (2014), pp. 312–336.

[7] B. Adcock and D. Huybrechs, *Frames and numerical approximation II: generalized sampling*, arXiv:1802.01950, (2018).

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

[9] B. Adcock and J. Ruan, *Parameter selection and numerical approximation properties of Fourier extensions from fixed data*, J. Comput. Phys., 273 (2014), pp. 453–471.

[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 (2011), pp. 6248–6270.

[11] P. Balazs and K. Gröchenig, *A guide to localized frames and applications to Galerkin-like representations of operators*, in Novel Methods in Harmonic Analysis with Applications to Numerical Analysis and Data Processing, I. Pesenson, Q. Le Gia, A. Mayeli, H. Mhaskar, and D.-X. Zhou, eds., Applied and Numerical Harmonic Analysis, Birkhäuser, 2017.

[12] J. J. Benedetto, *Irregular sampling and frames*, in Wavelets: A Tutorial in Theory and Applications, C. K. Chui, ed., Boca Raton, FL: CRC, 1994.

[13] D. Boffi, N. Cavallini, and L. Gastaldi, *The finite element immersed boundary method with distributed Lagrange multiplier*, SIAM J. Numer. Anal., 53 (2015), pp. 2584–2604.

[14] A. Böttcher, *Infinite matrices and projection methods*, in Lectures on operator theory and its applications (Waterloo, ON, 1994), vol. 3 of Fields Inst. Monogr., Amer. Math. Soc., Providence, RI, 1996, pp. 1–72.

[15] 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^∞*, Appl. Math. Comput., 161 (2005), pp. 591–597.

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

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

[18] 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 (2007), pp. 1094–1125.

[19] E. J. Candès and D. Donoho, *Ridgelets: a key to higher-dimensional intermittency?*, Phil. Trans. R. Soc. Lond. A, 357 (1999), pp. 2495–2509.

[20] E. J. Candès and D. Donoho, *New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities*, Comm. Pure Appl. Math., 57 (2004), pp. 219–266.

[21] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, *Compressed sensing with coherent and redundant dictionaries*, Appl. Comput. Harmon. Anal., 31 (2010), pp. 59–73.

[22] P. G. Casazza, *The art of frame theory*, Taiwanese J. Math., 4 (2000), pp. 129–202.

[23] P. G. Casazza and O. Christensen, *Approximation of the frame coefficients using finite-dimensional methods*, J. Electronic Imaging, 06 (1997), pp. 479–483.

[24] P. G. Casazza and O. Christensen, *Riesz frames and approximation of the frame coefficients*, Approx. Theory Appl., 14 (1998), pp. 1–11.

[25] P. G. Casazza and O. Christensen, *Approximation of the inverse frame operator and applications to gabor frames*, J. Approx. Theory, 103 (2000), pp. 338–356.

[26] P. G. Casazza and G. Kutyniok, eds., *Finite Frames: Theory and Applications*, Birkhäuser, 2013.

[27] Z. Chen and C.-W. Shu, *Recovering exponential accuracy from collocation point values of smooth functions with end-point singularities*, J. Comput. Appl. Math., 265 (2014), pp. 83–95.

[28] Z. Chen and C.-W. Shu, *Recovering exponential accuracy in fourier spectral methods involving piecewise smooth functions with unbounded derivative singularities*, Preprint, (2014).

[29] O. Christensen, *Frames and the projection method*, Appl. Comput. Harmon. Anal., 1 (1993), pp. 50–53.

[30] O. Christensen, *Frames containing a Riesz basis and approximation of the frame coefficients using finite dimensional methods*, J. Math. Anal. Appl., 199 (1996), pp. 256–270.

[31] O. Christensen, *Finite-dimensional approximation of the inverse frame operator and applications to Weyl–Heisenberg frames and wavelet frames*, J. Fourier Anal. Appl., 6 (2001), pp. 79–91.

[32] O. Christensen, *An Introduction to Frames and Riesz Bases*, Applied and Numerical Harmonic Analysis, Birkhäuser, 2nd ed., 2016.

[33] O. Christensen and A. Lindner, *Frames containing a Riesz basis and approximation of the inverse frame operator*, Internat. Ser. Numer. Math., 137 (2001), pp. 89–100.

[34] O. Christensen and T. Strohmer, *The finite section method and problems in frame theory*, J. Approx. Theory, 133 (2005), pp. 221–237.

[35] S. Dahlke, M. Fornasier, and T. Raasch, Advances in Computational Mathematics, 27 (2007), pp. 27–63.

[36] S. Datta, S. Howard, and D. Cochran, *Geometry of the Welch bounds*, Lin. Alg. Appl., 437 (2012), pp. 2455–2470.

[37] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.

[38] I. Daubechies, A. Grossmann, and Y. Meyer, *Painless nonorthogonal expansions*, J. Math. Phys., (1986), pp. 1271–1283.

[39] E. B. Davies and M. Plum, *Spectral pollution*, IMA J. Num. Anal., 23 (2004), pp. 417–438.

[40] M. DÖRFLER AND G. A. VELASCO, *Adaptive Gabor frames by projection onto time-frequency subspaces*, in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014.

[41] R. J. DUFFIN AND A. C. SCHAEFFER, *A class of nonharmonic Fourier series*, Trans. Amer. Math. Soc., 72 (1952), pp. 341–366.

[42] K. S. ECKHOFF, *Accurate and efficient reconstruction of discontinuous functions from truncated series expansions*, Math. Comp., 61 (1993), pp. 745–763.

[43] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer Academic, Dordrecht, The Netherlands, 1996.

[44] M. FICKUS, D. G. MIXON, AND J. C. TREMAIN, *Steiner equiangular tight frames*, Lin. Alg. Appl., 436 (2012), pp. 1014–1027.

[45] M. FORNASIER AND K. GRÖCHENIG, *Intrinsic localization of frames*, Constr. Approx., 22 (2005), pp. 395–415.

[46] K. GRÖCHENIG, *Localization of frames, Banach frames, and the invertibility of the frame operator*, J. Fourier Anal. Appl., 10 (2004), pp. 105–132.

[47] K. GRÖCHENIG, Z. RZESZOTNIK, AND T. STROHMER, *Quantitative estimates for the finite section method and Banach algebras of matrices*, Integral Equations and Operator Theory, 67 (2011), pp. 183–202.

[48] A. C. HANSEN, *On the approximation of spectra of linear operators on Hilbert spaces*, J. Funct. Anal., 254 (2008), pp. 2092–2126.

[49] A. C. HANSEN, *On the solvability complexity index, the n-pseudospectrum and approximations of spectra of operators*, J. Amer. Math. Soc., 24 (2011), pp. 81–124.

[50] P. C. HANSEN, V. PEREYRA, AND G. SCHERER, *Least Squares Data Fitting with Applications*, John Hopkins University Press, Baltimore, 2012.

[51] M. L. HARRISON, *Frames and irregular sampling from a computational perspective*, PhD thesis, University of Maryland – College Park, 1998.

[52] D. P. HEWETT, S. N. CHANDLER-WILDE, S. LANGDON, AND A. TWIGGER, *A high frequency boundary element method for scattering by a class of nonconvex obstacles*, Numer. Math., 129 (2015), pp. 647–689.

[53] J. A. HOGAN AND J. D. LAKEY, *Duration and Bandwidth Limiting*, Applied and Numerical Harmonic Analysis, Birkhäuser, 2012.

[54] D. HUYBRECHS, *On the Fourier extension of non-periodic functions*, SIAM J. Numer. Anal., 47 (2010), pp. 4326–4355.

[55] M. JAVED AND L. N. TREFETHEN, *Euler-Maclaurin and Gregory interpolants*, Numer Math, 132 (2016), pp. 201–216.

[56] F. KASOLIS, E. WADBRO, AND M. BERGGREN, *Analysis of fictitious domain approximations of hard scatterers*, SIAM J. Numer. Anal., 2015 (2015), pp. 2347–2362.

[57] J. KOVACEVIC AND A. CHEBIRA, *Life beyond bases: The advent of frames (part 2)*, IEEE Signal Process. Mag., 24 (2007), pp. 115–125.

[58] J. KOVACEVIC AND A. CHEBIRA, *Life beyond bases: The advent of frames (part i)*, IEEE Signal Process. Mag., 24 (2007), pp. 86–104.

[59] A. KRYLOV, *On approximate calculations*, Lectures delivered in 1906 (in Russian). St Petersburg, (1907).

[60] G. KUTYNIOK AND D. LABATE, eds., *Shearlets: Multiscale Analysis for Multivariate Data*, Springer, 2012.

[61] H. J. LANDAU AND H. O. POLLAK, *Prolate spheroidal wave functions, Fourier analysis and uncertainty—III: The dimension of the space of essentially timeand bandlimited signals*, Bell System Tech J., 41 (1962), pp. 1295–1336.

[62] M. LEWIN AND É. SÉRÉ, *Spectral pollution and how to avoid it*, Proc. London Math. Soc., 100 (2009), pp. 864–900.

[63] M. LINDNER, *Infinite Matrices and their Finite Sections*, Frontiers in Mathematics, Birkhäuser Verlag, Basel, 2006.

[64] S. H. LUI, *Spectral domain embedding for elliptic PDEs in complex domains*, J. Comput. Appl. Math., 225 (2009), pp. 541–557.

[65] M. LYON, *Approximation error in regularized SVD-based Fourier continuations*, Appl. Numer. Math., 62 (2012), pp. 1790–1803.

[66] M. LYON, *A fast algorithm for Fourier continuation*, SIAM J. Sci. Comput., 33 (2012), pp. 3241–3260.

[67] 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 (2010), pp. 3358–3381.

[68] S. G. MALLAT, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, 3 ed.,

2009.

[69] R. Matthysen and D. Huybrechs, *Fast algorithms for the computation of Fourier extensions of arbitrary length*, SIAM J. Sci. Comput., 38 (2016), pp. A899–A922.

[70] R. Matthysen and D. Huybrechs, *Function approximation on arbitrary domains using Fourier extension frames*, arXiv:1706.04848, (2017).

[71] A. Neumaier, *Solving ill-conditioned and singular linear systems: a tutorial on regularization*, SIAM Rev., 40 (1998), pp. 636–666.

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

[73] R. Platte, A. J. Gutierrez, and A. Gelb, *Edge informed Fourier reconstruction from non-uniform spectral data with exponential convergence rates*, Preprint, (2012).

[74] J. M. Renes, *Equiangular tight frames from Paley tournaments*, Lin. Alg. Appl., 426 (2007), pp. 497–501.

[75] D. Shirokoff and J.-C. Nave, *A sharp-interface active penalty method for the incompressible Navier–Stokes equations*, J. Sci. Comput., 62 (2015), pp. 53–77.

[76] D. Slepian, *Prolate spheriodal wave functions. Fourier analysis, and uncertainty V: The discrete case*, Bell System Tech J., 57 (1978), pp. 1371–1430.

[77] G. Song, J. Davis, and A. Gelb, *A high-dimensional inverse frame operator approximation technique*, SIAM J. Numer. Anal., 54 (2016), pp. 2282–2301.

[78] G. Song and A. Gelb, *Approximating the inverse frame operator from localized frames*, Appl. Comput. Harmon. Anal., 35 (2013), pp. 94–110.

[79] D. B. Stein, R. D. Guy, and B. Thomases, *Immersed boundary smooth extension: a high-order method for solving PDE on arbitrary smooth domains using Fourier spectral methods*, J. Comput. Phys., 304 (2016), pp. 252–274.

[80] R. Stevenson, *Adaptive solution of operator equations using wavelet frames*, SIAM J. Numer. Anal., 41 (2003), pp. 1074–1100.

[81] T. Strohmer, *Numerical analysis of the nonuniform sampling problem*, J. Comput. Appl. Math., 122 (2000), pp. 297–316.

[82] T. Strohmer, *A note on equiangular tight frames*, Lin. Alg. Appl., 429 (2008), pp. 326–330.

[83] M. A. Sustik, J. A. Tropp, I. S. Dhillon, and R. W. Heath, *On the existence of equiangular tight frames*, Lin. Alg. Appl., 426 (2007), pp. 619–635.

[84] A. Teolis and J. J. Benedetto, *Local frames and noise reduction*, Signal Process., 45 (1995), pp. 369–387.

[85] L. N. Trefethen, *Approximation Theory and Approximation Practice*, SIAM, Philadelphia, 2013.