Representation and Compression of Multi-Dimensional Piecewise Functions Using *Surflets*

Venkat Chandrasekaran,[m] *Michael B. Wakin,*[r] *Dror Baron,*[r] *Richard G. Baraniuk*[r]

[m] Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

[r] Department of Electrical and Computer Engineering
Rice University

March 21, 2006

Abstract

We study the representation, approximation, and compression of functions in M dimensions that consist of constant or smooth regions separated by smooth $(M-1)$-dimensional discontinuities. Examples include images containing edges, video sequences of moving objects, and seismic data containing geological horizons. For both function classes, we derive the optimal asymptotic approximation and compression rates based on Kolmogorov metric entropy. For piecewise constant functions, we develop a multiresolution predictive coder that achieves the optimal rate-distortion performance; for piecewise smooth functions, our coder has near-optimal rate-distortion performance. Our coder for piecewise constant functions employs *surflets*, a new multiscale geometric tiling consisting of M-dimensional piecewise constant atoms containing polynomial discontinuities. Our coder for piecewise smooth functions uses *surfprints*, which wed surflets to wavelets for piecewise smooth approximation. Both of these schemes achieve the optimal asymptotic approximation performance. Key features of our algorithms are that they carefully control the potential growth in surflet parameters at higher smoothness and do not require explicit estimation of the discontinuity. We also extend our results to the corresponding discrete function spaces for sampled data. We provide asymptotic performance results for both discrete function spaces and relate this asymptotic performance to the sampling rate and smoothness orders of the underlying functions and discontinuities. For approximation of discrete data we propose a new scale-adaptive dictionary that contains few elements at coarse and fine scales, but many elements at medium scales. Simulation results demonstrate that surflets provide superior compression performance when compared to other state-of-the-art approximation schemes.

This work was supported by NSF grant CCF–0431150, ONR grant N00014–02–1–0353, AFOSR grant FA9550–04–0148, AFRL grant FA8650–051850, and the Texas Instruments Leadership University Program. Preliminary versions of this work have appeared at the 38th Annual Conference on Information Sciences and Systems [1] and the 2004 IEEE International Symposium on Information Theory [2]. Email: venkatc@mit.edu and {wakin, drorb, richb}@rice.edu; Web: dsp.rice.edu

1 Introduction

1.1 Motivation

Sparse signal representations feature prominently in a broad variety of signal processing applications. The advantages of characterizing signals using just a few elements from a dictionary are clear in the case of compression [3–6], but also extend to tasks such as estimation [7–10] and classification [11, 12]. The *dimensionality* of signals is an important factor in determining sparse representations. Wavelets provide an appropriate representation for smooth one-dimensional (1D) signals, achieving the optimal asymptotic representation behavior for this class [13, 14], and they maintain this optimal performance even for smooth 1D signals containing a finite number of point discontinuities.

Unfortunately, this optimality does not extend completely to 2D signals due to the different nature of discontinuities in two dimensions [15, 16]. While smooth signals in 2D containing a finite number of *point* singularities are sparsely represented by a wavelet basis, 2D piecewise smooth signals containing discontinuities along 1D curves ("edges") are not represented efficiently by a wavelet basis.

The problem is that the isotropically scaled wavelet basis fails to capture the anisotropic structure of 1D discontinuities. Due to its global support, the Fourier basis also fails to effectively capture the local nature of these discontinuities. Nonetheless, 2D piecewise smooth signals containing 1D discontinuities are worthy candidates for sparse representation. The 1D discontinuities often carry interesting information, since they signify a boundary between two regions. Edges in images illustrate this point well. Lying along 1D curves, edges provide fundamental information about the geometry of a scene. Therefore, any signal processing application that relies on a sparse representation requires an efficient tool for representing discontinuities.

A growing awareness of the limitations of traditional bases for representing 2D signals with 1D discontinuities has resulted in new multiscale representations. (Multiscale representations offer a number of advantages for signal processing, enabling, for example, predictive and progressive source coding, powerful statistical models, and efficient tree-structured processing.) The resulting solutions fall into two broad categories: tight frames and geometric tilings. Loosely speaking, *tight frames* refer to dictionaries from which approximations are formed using linear combinations of atomic functions, while *geometric tiling* schemes create an adaptive partitioning of the signal to which local approximation elements are assigned. The key factor exploited by all these solutions is that the discontinuities often have a coherent geometric structure in one dimension. A primary solution proposed in the class of 2D tight frames is the curvelet dictionary [16]. The salient feature of curvelets is an anisotropic scaling of the atomic elements, with increased directionality at finer resolution; the resulting dictionary is well-suited to represent 1D singularities. In the category of geometric tilings, wedgelets [8] are piecewise constant functions defined on 2D dyadic squares, where a linear edge singularity separates the two constant regions. Tilings of wedgelets at various scales combine to form piecewise linear approximations to the 1D edges in images. Each of these solutions assumes that the underlying 1D discontinuity is $\mathcal{C}^2$-smooth. The class of 2D signals containing such smooth 1D discontinuities is often referred to as the *Horizon* function model [8].

Unfortunately, the tight frames and geometric tilings proposed to date also face certain limitations. First, none of these techniques is directly applicable to problems in higher dimensions beyond 2D. Signals in video (3D), geophysics (3D, 4D), and light-field imaging (4D, 5D) [17] frequently

contain information-rich discontinuities separating different regions. The confounding aspect, as before, is that these discontinuities often have a well-defined geometric structure in one lower dimension than that of the signal. In video signals, for example, where we can imagine stacking a 2D image sequence to form a 3D volume, 3D regions are separated by smooth, well-structured 2D discontinuities traced out by the moving boundaries of objects in the video. Therefore, we can model discontinuities in M-dimensional signals as smooth functions of $M-1$ variables. This is an analogous extension to the *Horizon* function model [8] used for 2D signals with 1D smooth discontinuities.

Another limitation of current frame and tiling approaches is that they are intended primarily for signals with underlying $\mathcal{C}^2$-smooth discontinuities. There exist signals, however, for which the discontinuities are inherently of a higher order of smoothness [18]. While $\mathcal{C}^K \subset \mathcal{C}^2$ for $K > 2$, dictionaries achieving the optimal performance can be constructed only if the full smoothness of the discontinuities is exploited. Interesting mathematical insights can also be obtained by considering higher orders of smoothness. Finally, some of the proposed solutions such as wedgelets [8] consider representation of piecewise constant signals in higher dimensions. Real-world multi-dimensional signals, however, often consist of discontinuities separating smooth (but not constant) regions. This motivates a search for sparse representations for *signals in any dimension* consisting of *regions of arbitrary smoothness* that are separated by *discontinuities in one lower dimension* of *arbitrary smoothness*.

1.2 Approximation and Compression

In this paper, we address the problem of approximating and compressing M-dimensional signals that contain a smooth $(M-1)$-dimensional discontinuity separating regions in M dimensions[1] (see Fig. 1 for examples in 2D and 3D). The discontinuities in our models have smoothness $\mathcal{C}^{K_\mathrm{d}}$ in $M-1$ dimensions. They separate two regions that may be constant (Sec. 1.3.1) or $\mathcal{C}^{K_\mathrm{s}}$-smooth in M dimensions (Sec. 1.3.2). Our *approximation results* characterize the number of atoms required to efficiently represent a signal and are useful for tasks such as estimation [8, 10] and classification [11, 12]. The measure for approximation performance is the asymptotic rate of decay of the distortion between a signal and its approximant as the number of atoms becomes large. *Compression results*, on the other hand, also take into account *which* atoms are used in constructing a sparse approximation and are crucial for communicating an approximation to a signal. Compression performance is assessed by considering the asymptotic rate-distortion behavior, which specifies the rate of decay of the distortion between a signal and its approximant as the number of bits used to encode the approximant becomes large. Since it is impossible to determine for abstract function classes precise approximation or rate-distortion results, we will state our performance results in terms of *metric entropy* [19], which characterizes the order of the number of atoms or bits required to describe a function in the class under consideration.

1.3 Contributions

We consider two function classes as models for M-dimensional signals containing $(M-1)$-dimensional discontinuities. This section outlines our contributions toward constructing efficient representations for elements of these classes, including extensions to discrete data.

[1] We discuss possible extensions to signals containing multiple discontinuities in Sec. 4.6.

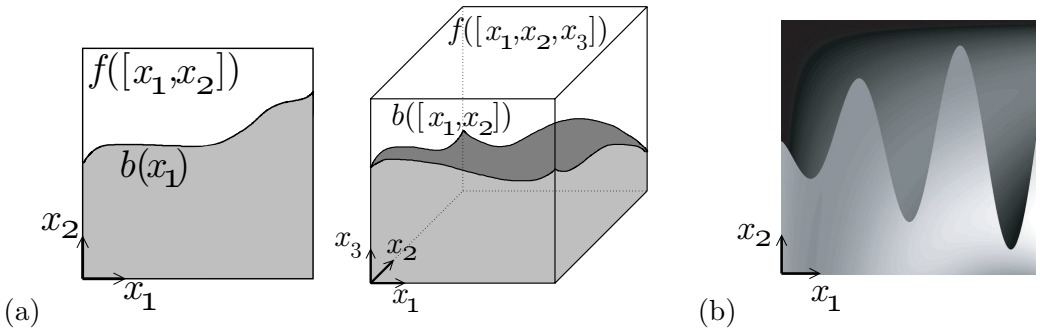


Figure 1: *(a) Piecewise constant ("Horizon-class") functions for dimensions $M = 2$ and $M = 3$. (b) Piecewise smooth function for dimension $M = 2$.*

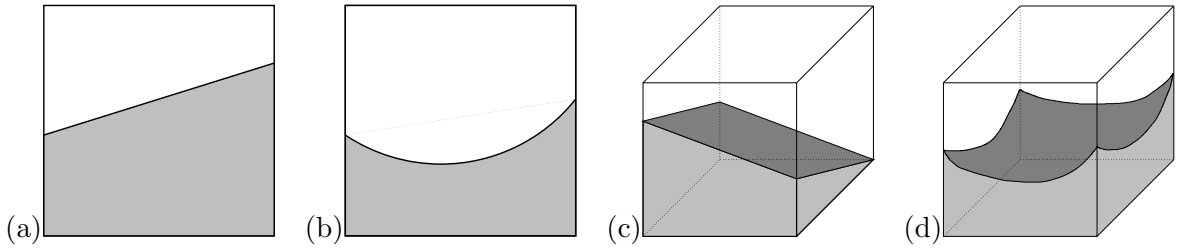


Figure 2: *Example surflets, designed for (a) $M = 2$, $K_{\mathrm{d}} \in (1, 2]$; (b) $M = 2$, $K_{\mathrm{d}} \in (2, 3]$; (c) $M = 3$, $K_{\mathrm{d}} \in (1, 2]$; (d) $M = 3$, $K_{\mathrm{d}} \in (2, 3]$.*

1.3.1 Piecewise constant M-dimensional functions

We consider the function class $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$, consisting of M-dimensional piecewise constant Horizon functions [8] that contain a $\mathcal{C}^{K_{\mathrm{d}}}$-smooth $(M-1)$-dimensional discontinuity separating two *constant* regions. We begin by deriving the optimal asymptotic rates for nonlinear approximation and compression of functions in the class $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$. These bounds extend the results of Cohen et al. [14], Clements [20], and Kolmogorov and Tihomirov [19], which characterize the optimal asymptotic approximation and compression behavior of $(M-1)$-dimensional smooth functions.

We introduce a new M-dimensional geometric tiling framework for multiscale representation of functions in $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$. We propose a dictionary of atoms defined on dyadic hypercubes, where each atom is an M-dimensional piecewise constant function containing an $(M-1)$-dimensional polynomial discontinuity. We christen these atoms *surflets* after Richard Surflet, a 16^{th} century translator, occultist, and "practitioner in physicke" [21], and also after the fact that they resemble small pieces of surfaces in higher dimensional space (see Fig. 2 for examples in 2D and 3D).

The surflet dictionary is a generalization of the wedgelet dictionary [8] to higher-dimensional signals containing discontinuities of arbitrary smoothness (a wedgelet is a surflet with $K_{\mathrm{d}} = 2$ and $M = 2$). We show that tilings of elements drawn from the the surflet dictionary can achieve the optimal approximation rate

$$\left\| f^{\mathrm{c}} - \widehat{f^{\mathrm{c}}_N} \right\|^2_{L_2} \lesssim \left(\frac{1}{N} \right)^{\frac{K_{\mathrm{d}}}{M-1}}$$

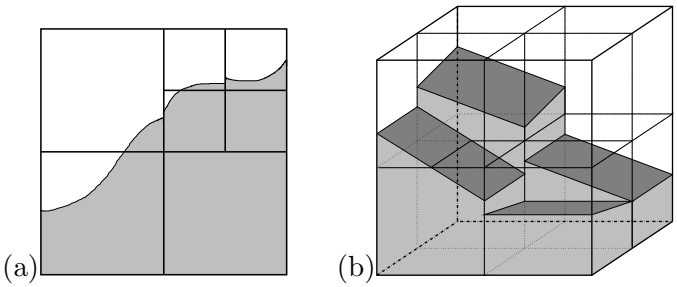


Figure 3: *Example surflet tilings, (a) piecewise cubic with $M = 2$ and (b) piecewise linear with $M = 3$.*

for the class $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$,[2] where $f^{\mathrm{c}} \in \mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ and $\widehat{f_N^{\mathrm{c}}}$ is the best N-term approximant to f^{c}. Example 2D and 3D surflet tilings appear in Fig. 3.

We also propose a tree-structured compression algorithm for M-dimensional functions using surflets and establish that this algorithm achieves the optimal rate-distortion performance

$$\left\| f^{\mathrm{c}} - \widehat{f_R^{\mathrm{c}}} \right\|_{L_2}^2 \lesssim \left(\frac{1}{R} \right)^{\frac{K_{\mathrm{d}}}{M-1}}$$

for the class $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$. Here, $\widehat{f_R^{\mathrm{c}}}$ is the best approximation to f^{c} that can be encoded using R bits.

Our approach incorporates the following major features:

- the ability to operate directly on the M-dimensional function, without requiring explicit knowledge (or estimation) of the $(M-1)$-dimensional discontinuity;
- the use of multiresolution predictive coding in order to realize significant gains in rate-distortion performance; and
- a technique to quantize and encode higher-order polynomial coefficients with lesser precision without a substantial increase in distortion.

Without such a quantization scheme, higher-order polynomials would be impractical for representing boundaries smoother than $\mathcal{C}^2$, due to an exponential explosion in the number of polynomial parameters and thus the size of the dictionary. A fascinating aspect of our solution is that the size of the surflet dictionary can be reduced tremendously without sacrificing the approximation capability.

1.3.2 Piecewise smooth M-dimensional functions

We consider the function class $\mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$, consisting of M-dimensional piecewise *smooth* functions that contain a $\mathcal{C}^{K_{\mathrm{d}}}$-smooth $(M-1)$-dimensional discontinuity separating two regions that are $\mathcal{C}^{K_{\mathrm{s}}}$-smooth in M dimensions. We establish the optimal approximation and rate-distortion bounds for this class.

[2] We focus here on *asymptotic* performance. We use the notation $f(\alpha) \lesssim g(\alpha)$, or $f(\alpha) = O(g(\alpha))$, if there exists a constant C, possibly large but not dependent on the argument α, such that $f(\alpha) \leq Cg(\alpha)$.

Despite their ability to efficiently describe a discontinuity, surflets alone are not a suitable representation for functions in $\mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$ because a piecewise *constant* approximation is inefficient on the smooth M-dimensional regions. Conversely, we note that wavelets are well-suited for representing these M-dimensional smooth regions, yet they fail to efficiently capture discontinuities. Thus, we propose a dictionary consisting of the following elements:

- an M-dimensional basis of compactly supported *wavelet* atoms having sufficiently many vanishing moments [22]; and
- a dictionary of M-dimensional *surfprint* atoms. Each surfprint can be viewed as a linear combination of wavelet basis functions. It is derived by projecting a piecewise *polynomial* surflet atom (having two M-dimensional polynomial regions separated by a polynomial discontinuity) onto a wavelet subspace.

The surfprint dictionary is an extension of the wedgeprint dictionary [23] for higher-dimensional signals having regions of arbitrary smoothness separated by discontinuities of arbitrary smoothness (a wedgeprint is a surfprint with $K_{\mathrm{d}} = 2$, $K_{\mathrm{s}} = 2$, and $M = 2$). Constructed as projections of piecewise polynomial surflets onto wavelet subspaces, surfprints interface naturally with wavelets for sparse approximation of piecewise smooth M-dimensional functions. We show that the combined wavelet/surfprint dictionary achieves the optimal approximation rate for $\mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$:

$$\left\| f^{\mathrm{s}} - \widehat{f^{\mathrm{s}}_N} \right\|_{L_2}^2 \lesssim \left(\frac{1}{N} \right)^{\min\left(\frac{K_{\mathrm{d}}}{M-1}, \frac{2K_{\mathrm{s}}}{M} \right)}.$$

Here, $f^{\mathrm{s}} \in \mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$, and $\widehat{f^{\mathrm{s}}_N}$ is the best N-term approximant to f^{s}. We include a careful treatment of the surfprint polynomial degrees and the number of wavelet vanishing moments required to attain the optimal approximation rates. We also propose a tree-based encoding scheme that comes within a logarithmic factor of the optimal rate-distortion performance for this class:

$$\left\| f^{\mathrm{s}} - \widehat{f^{\mathrm{s}}_R} \right\|_{L_2}^2 \lesssim \left(\frac{\log R}{R} \right)^{\min\left(\frac{K_{\mathrm{d}}}{M-1}, \frac{2K_{\mathrm{s}}}{M} \right)}.$$

Here, $\widehat{f^{\mathrm{s}}_R}$ is the best approximation to f^{s} that can be encoded using R bits.

1.3.3 Extensions to discrete data

We also address the problem of representing discrete data obtained by sampling a continuous function from $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ or $\mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$. We denote these classes of discretized (or "voxelized") data by $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$ and $\widetilde{\mathcal{F}_{\mathrm{S}}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$, respectively, and we allow for different sampling rates in each of the M dimensions.

In order to efficiently represent data from $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$, we use a dictionary of discrete surflet atoms derived by voxelizing the surflet atoms of our continuous surflet dictionary. We show that, up to a certain critical scale (which depends on K_{d} and the sampling rates in each dimension), the approximation results for $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ extend to $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$. Beyond this scale, however, voxelization effects dominate, and the surflets designed for elements of $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ offer unnecessary precision for representing elements of $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$. To account for these effects, we propose a new

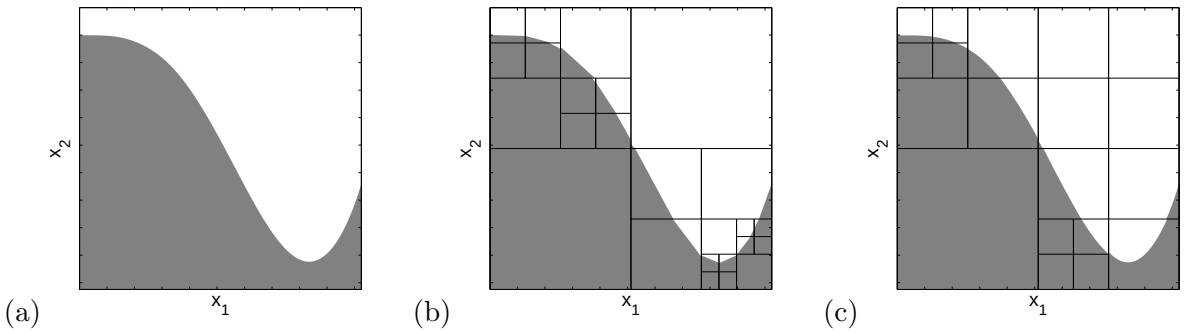


Figure 4: *Comparison of pruned surflet tilings using two surflet dictionaries (see Sec. 6 for more details). (a) Test image with $M = 2$ and $K_{\mathrm{d}} = 3$. (b) The wedgelets from Dictionary 2 can be encoded using* 482 *bits and yields PSNR 29.86dB. (c) The quadratic/wedgelet combination from Dictionary 3 can be encoded using only* 275 *bits and yields PSNR 30.19dB.*

scale-adaptive dictionary that contains few elements at coarse and fine scales, but many elements at medium scales. We also present simulation results confirming the advantages of our surflet coding strategies; see Fig. 4 for an example demonstrating the benefits of quadratic surflets.[3]

For elements of $\widetilde{\mathcal{F}_{\mathrm{S}}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$, we use a dictionary of compactly supported discrete wavelet basis functions. The number of discrete vanishing moments required is the same as that of the continuous wavelet basis applied to members of $\mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$. Discrete surfprint atoms are obtained by projecting discrete surflet atoms onto this discrete wavelet basis. As before, we see a critical scale at which voxelization effects begin to dominate; again these effects can be addressed by the appropriate scale-adaptive surfprint dictionary.

1.4 Relation to previous work

Our work can be viewed as a generalization of wedgelet [8] and wedgeprint [23] representations. Our extensions, however, provide fundamental new insights in the following directions:

- The wedgelet and wedgeprint dictionaries are restricted to 2D signals, while our proposed representations are relevant in higher dimensions.
- Wedgelets and wedgeprints achieve optimal approximation rates only for functions that are $\mathcal{C}^2$-smooth and contain a $\mathcal{C}^2$-smooth discontinuity; our results not only show that surflets and surfprints can be used to achieve optimal rates for more general classes, but also highlight the necessary polynomial quantization scheme (a nontrivial extension from wedgelets).
- In the construction of surfprint atoms, we derive the surfprint polynomial degrees required for optimal performance as a function of K_{s}, K_{d}, and M. Such insight cannot be gained from wedgeprints, which are derived simply from the projection of piecewise constant wedgelets.
- We also present a more thorough analysis of discretization effects, including new insights on the multiscale behavior (not revealed by considering wedgelets alone), a new strategy for re-

[3]The Peak Signal-to-Noise Ratio (PSNR) we quote is a common measure of distortion that derives from the mean-square error (MSE); assuming a maximum possible signal intensity of I, $PSNR := 10 \log_{10} \frac{I^2}{MSE}$.

ducing the surflet dictionary size at fine scales, and the first treatment of wedgeprint/surfprint discretization.

During final preparation of this manuscript we learned of a related generalization of wedgelets to quadratic polynomials [24]. Novel components of our work, however, include its broad generality to arbitrary orders of smoothness and arbitrary dimension; the quantization scheme, predictive coding, and analytical results; the scale-adaptive dictionary to deal with voxelization; and the surfprints construction.

1.5 Paper organization

In Sec. 2, we define our function models and state the specific goals of our approximation and compression algorithms. We introduce surflets in Sec. 3. In Sec. 4, we describe our surflet-based representation schemes for elements of $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ in detail. In Sec. 5, we present our novel dictionary of wavelets and surfprints for effectively representing elements of $\mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$. Section 6 discusses extensions to discrete data from $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$ and $\widetilde{\mathcal{F}_{\mathrm{S}}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$. Section 7 summarizes our contributions and insights. The appendices provide additional details and proofs of all the theorems.

2 Background

2.1 Lipschitz smoothness

A function of D variables has smoothness of order $K > 0$, where $K = r + \alpha$, r is an integer, and $\alpha \in (0, 1]$, if the following criteria are met [19, 20]:

- All iterated partial derivatives with respect to the D directions up to order r exist and are continuous.
- All such partial derivatives of order r satisfy a Lipschitz condition of order α (also known as a Hölder condition).[4]

We consider the space of smooth functions whose partial derivatives up to order r are bounded by some constant Ω. We denote the space of such bounded functions with bounded partial derivatives by $\mathcal{C}^K$, where this notation carries an implicit dependence on Ω. Observe that $r = \lceil K - 1 \rceil$, where $\lceil \cdot \rceil$ denotes rounding up. Also, when K is an integer $\mathcal{C}^K$ includes as a subset the traditional space "C^K" (the class of functions that have $K = r + 1$ continuous partial derivatives).

2.2 Multi-dimensional signal model

Let $\mathbf{x} \in [0, 1]^M$, and let x_i denote its i'th element, where boldface characters are used to denote vectors. We denote the first $M - 1$ elements of $\mathbf{x}$ by $\mathbf{y}$, i.e., $\mathbf{y} = [x_1, x_2, \cdots, x_{M-1}] \in [0, 1]^{M-1}$. Let f, g_1, and g_2 be functions of M variables

$$f, g_1, g_2 : [0, 1]^M \to \mathbb{R},$$

[4] A function $d \in$ Lip(α) if $|d(\mathbf{z}_1 + \mathbf{z}_2) - d(\mathbf{z}_1)| \leq C\|\mathbf{z}_2\|^\alpha$ for all D-dimensional vectors $\mathbf{z}_1, \mathbf{z}_2$.

and let b be a function of $M-1$ variables

$$b : [0,1]^{M-1} \to \mathbb{R}.$$

We define the function f in the following piecewise manner:

$$f(\mathbf{x}) = \begin{cases} g_1(\mathbf{x}), & x_M \geq b(\mathbf{y}) \\ g_2(\mathbf{x}), & x_M < b(\mathbf{y}). \end{cases}$$

2.2.1 Piecewise constant model

We begin by considering the "piecewise constant" case when $g_1 = 1$ and $g_2 = 0$. The $(M-1)$-dimensional discontinuity b defines a boundary between two constant regions in M dimensions. The piecewise constant functions f defined in this manner are Horizon-class functions [8]. When $b \in \mathcal{C}^{K_\mathrm{d}}$, with $K_\mathrm{d} = r_\mathrm{d} + \alpha_\mathrm{d}$, we denote the resulting space of functions f by $\mathcal{F}_\mathrm{C}(M, K_\mathrm{d})$. When $M = 2$, these functions can be interpreted as images containing a $\mathcal{C}^{K_\mathrm{d}}$-smooth one-dimensional discontinuity that separates a 0-valued region below from a 1-valued region above. For $M = 3$, functions in $\mathcal{F}_\mathrm{C}(M, K_\mathrm{d})$ can be interpreted as cubes with a 2D $\mathcal{C}^{K_\mathrm{d}}$-smooth surface cutting through them, dividing them into two regions — 0-valued below the surface and 1-valued above it (see Fig. 1(a) for examples in 2D and 3D).

2.2.2 Piecewise smooth model

Next, we define a model for piecewise smooth functions. For this class of functions, we let $g_1, g_2 \in \mathcal{C}^{K_\mathrm{s}}$, with $K_\mathrm{s} = r_\mathrm{s} + \alpha_\mathrm{s}$, and $b \in K_\mathrm{d}$, with $K_\mathrm{d} = r_\mathrm{d} + \alpha_\mathrm{d}$. The resulting piecewise smooth function f consists of an $(M-1)$-dimensional $\mathcal{C}^{K_\mathrm{d}}$-smooth discontinuity that separates two $\mathcal{C}^{K_\mathrm{s}}$-smooth regions in M dimensions (see Fig. 1(b) for an example in 2D). We denote the class of such piecewise smooth functions by $\mathcal{F}_\mathrm{S}(M, K_\mathrm{d}, K_\mathrm{s})$. One can check that both $\mathcal{F}_\mathrm{C}(M, K_\mathrm{d})$ and the space of M-dimensional uniformly $\mathcal{C}^{K_\mathrm{s}}$ functions are subsets of $\mathcal{F}_\mathrm{S}(M, K_\mathrm{d}, K_\mathrm{s})$.

2.3 Approximation and compression performance measures

In this paper, we define dictionaries of atoms from which we construct an approximation $\widehat{f}$ to f, which may belong to $\mathcal{F}_\mathrm{C}(M, K_\mathrm{d})$ or $\mathcal{F}_\mathrm{S}(M, K_\mathrm{d}, K_\mathrm{s})$. We analyze the performance of our coding scheme using the L_p distortion measure between the M-dimensional functions f and $\widehat{f}$

$$\left\| f - \widehat{f} \right\|^p_{L_p([0,1]^M)} = \int_{[0,1]^M} |f - \widehat{f}|^p$$

with $p = 2$ (i.e., the standard squared-L_2 distortion measure). We measure the ability of our dictionary of atoms to represent f sparsely by the *asymptotic approximation performance* $\|f - \widehat{f_N}\|^2_{L_2}$. Here, N is the number of atomic elements from the dictionary used to construct an approximation to f and $\widehat{f_N}$ is the best N-term approximant to f. We consider the rate of decay of $\|f - \widehat{f_N}\|^2_{L_2}$ as $N \to \infty$.

We also present compression algorithms that encode those atoms from the corresponding dictionaries (depending on whether $f \in \mathcal{F}_\mathrm{C}(M, K_\mathrm{d})$ or $f \in \mathcal{F}_\mathrm{S}(M, K_\mathrm{d}, K_\mathrm{s})$) used to construct $\widehat{f}$. We

measure the performance of these compression algorithms by the *asymptotic rate-distortion function* $\|f - \widehat{f_R}\|_{L_2}^2$, where $\widehat{f_R}$ is the best approximation to f that can be encoded using R bits [25]. We consider the behavior of $\|f - \widehat{f_R}\|_{L_2}^2$ as $R \to \infty$.

In [14], Cohen et al. establish the optimal approximation rate for D-dimensional $\mathcal{C}^K$-smooth functions d:

$$\left\|d - \widehat{d_N}\right\|_{L_2}^2 \lesssim \left(\frac{1}{N}\right)^{\frac{K}{D}}.$$

Similarly, the results of Clements [20] (extending those of Kolmogorov and Tihomirov [19]) regarding metric entropy establish bounds on the optimal achievable asymptotic rate-distortion performance for D-dimensional $\mathcal{C}^K$-smooth functions d:

$$\left\|d - \widehat{d_R}\right\|_{L_2}^2 \lesssim \left(\frac{1}{R}\right)^{\frac{K}{D}}.$$

These results, however, are only useful for characterizing optimal *separate* representations for the $(M-1)$-dimensional discontinuity ($D = M-1,\ K = K_{\mathrm{d}},\ d = b$ in Sec. 2.1) and the M-dimensional smooth regions ($D = M,\ K = K_{\mathrm{s}},\ d = g_1, g_2$ in Sec. 2.1).

We extend these results to non-separable representations of the M-dimensional function classes $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ and $\mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$ in Theorems 1 and 2, respectively.

Theorem 1 *The optimal asymptotic approximation performance that can be obtained for all $f^{\mathrm{c}} \in \mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ is given by*

$$\left\|f^{\mathrm{c}} - \widehat{f_N^{\mathrm{c}}}\right\|_{L_2}^2 \lesssim \left(\frac{1}{N}\right)^{\frac{K_{\mathrm{d}}}{M-1}}.$$

Similarly, the optimal asymptotic compression performance that can be obtained for all $f^{\mathrm{c}} \in \mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ is given by

$$\left\|f^{\mathrm{c}} - \widehat{f_R^{\mathrm{c}}}\right\|_{L_2}^2 \lesssim \left(\frac{1}{R}\right)^{\frac{K_{\mathrm{d}}}{M-1}}.$$

Implicit in the proof of the above theorem, which appears in Appendix A, is that any scheme that is optimal for representing and compressing the M-dimensional function $f^{\mathrm{c}} \in \mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ in the squared-L_2 sense is equivalently optimal for the $(M-1)$-dimensional discontinuity in the L_1 sense. Roughly, the squared-L_2 distance between two Horizon-class functions f_1^{c} and f_2^{c} over an M-dimensional domain $\mathcal{D} = [\mathcal{D}_{\mathrm{b}}^1, \mathcal{D}_{\mathrm{e}}^1] \times \cdots \times [\mathcal{D}_{\mathrm{b}}^M, \mathcal{D}_{\mathrm{e}}^M]$ is equal to the L_1 distance over the $(M-1)$-dimensional subdomain $[\mathcal{D}_{\mathrm{b}}^1, \mathcal{D}_{\mathrm{e}}^1] \times \cdots \times [\mathcal{D}_{\mathrm{b}}^{M-1}, \mathcal{D}_{\mathrm{e}}^{M-1}]$ between the $(M-1)$-dimensional discontinuities b_1^{c} and b_2^{c} in f_1^{c} and f_2^{c} respectively.

More precisely and for future reference, for every $\mathbf{y}$ in the $(M-1)$-dimensional subdomain of $\mathcal{D}$, we define the $\mathcal{D}$-*clipping* of an $(M-1)$-dimensional function b as

$$\overline{b}(\mathbf{y}) = \begin{cases} b(\mathbf{y}), & \mathcal{D}_{\mathrm{b}}^M \le b(\mathbf{y}) \le \mathcal{D}_{\mathrm{e}}^M \\ \mathcal{D}_{\mathrm{e}}^M, & b(\mathbf{y}) > \mathcal{D}_{\mathrm{e}}^M \\ \mathcal{D}_{\mathrm{b}}^M, & b(\mathbf{y}) < \mathcal{D}_{\mathrm{b}}^M. \end{cases}$$

The $\mathcal{D}$-*active region* of b is defined to be $\left\{\mathbf{y} \in [\mathcal{D}_{\mathrm{b}}^1, \mathcal{D}_{\mathrm{e}}^1] \times \cdots \times [\mathcal{D}_{\mathrm{b}}^{M-1}, \mathcal{D}_{\mathrm{e}}^{M-1}] : b(\mathbf{y}) \in [\mathcal{D}_{\mathrm{b}}^M, \mathcal{D}_{\mathrm{e}}^M]\right\}$, that subset of the subdomain of $\mathcal{D}$ for which the range of b lies in $[\mathcal{D}_{\mathrm{b}}^M, \mathcal{D}_{\mathrm{e}}^M]$. The $\mathcal{D}$-*clipped* L_1 *distance* between b_1^{c} and b_2^{c} is then defined as

$$\overline{L_1}(b_1^{\mathrm{c}}, b_2^{\mathrm{c}}) = \left\|\overline{b_1^{\mathrm{c}}} - \overline{b_2^{\mathrm{c}}}\right\|_{L_1([\mathcal{D}_{\mathrm{b}}^1, \mathcal{D}_{\mathrm{e}}^1] \times \cdots \times [\mathcal{D}_{\mathrm{b}}^{M-1}, \mathcal{D}_{\mathrm{e}}^{M-1}])}.$$

One can check that $\|f_1^{\mathrm{c}} - f_2^{\mathrm{c}}\|_{L_2(\mathcal{D})}^2 = \overline{L_1}(b_1^{\mathrm{c}}, b_2^{\mathrm{c}})$ for any $\mathcal{D}$.

The following theorem, which is proved in Appendix B, characterizes the optimal achievable asymptotic approximation rate and rate-distortion performance for approximating and encoding elements of the function class $\mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$.

Theorem 2 *The optimal asymptotic approximation performance that can be obtained for all* $f^{\mathrm{s}} \in \mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$ *is given by*

$$\left\|f^{\mathrm{s}} - \widehat{f_N^{\mathrm{s}}}\right\|_{L_2}^2 \lesssim \left(\frac{1}{N}\right)^{\min\left(\frac{K_{\mathrm{d}}}{M-1}, \frac{2K_{\mathrm{s}}}{M}\right)}.$$

Similarly, the optimal asymptotic compression performance that can be obtained for all $f^{\mathrm{s}} \in \mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$ *is given by*

$$\left\|f^{\mathrm{s}} - \widehat{f_R^{\mathrm{s}}}\right\|_{L_2}^2 \lesssim \left(\frac{1}{R}\right)^{\min\left(\frac{K_{\mathrm{d}}}{M-1}, \frac{2K_{\mathrm{s}}}{M}\right)}.$$

2.4 "Oracle" coders and their limitations

In order to approximate or compress an arbitrary function $f^{\mathrm{c}} \in \mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$, an algorithm is given the function f^{c}; we denote its $(M-1)$-dimensional $\mathcal{C}^{K_{\mathrm{d}}}$-smooth discontinuity by b^{c}. As constructed in Sec. 2.2, all of the critical information about f^{c} is contained in the discontinuity b^{c}. One would expect any efficient coder to exploit such a fact; methods through which this is achieved may vary.

One can imagine a coder that *explicitly* encodes an approximation $\widehat{b^{\mathrm{c}}}$ to b^{c} and then constructs a Horizon approximation $\widehat{f^{\mathrm{c}}}$. Knowledge of b^{c} could be provided from an external "oracle" [26], or b^{c} could conceivably be estimated from the provided data f^{c}. Wavelets provide an efficient method for compressing the $(M-1)$-dimensional smooth function b^{c}. Cohen et al. [14] describe a tree-structured wavelet coder that can be used to compress b^{c} with optimal rate-distortion performance in the L_1 sense. It follows that this wavelet coder is optimal (in the squared-L_2 sense) for coding instances of f^{c} at the optimal rate of Theorem 1. In practice, however, a coder is not provided with explicit information of b^{c}, and a method for estimating b^{c} from f^{c} may be difficult to implement. Estimates for b^{c} may also be quite sensitive to noise in the data.

A similar strategy could also be employed for $f^{\mathrm{s}} \in \mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$. For such a function, we denote the $(M-1)$-dimensional $\mathcal{C}^{K_{\mathrm{d}}}$-smooth discontinuity by b^{s} and the M-dimensional $\mathcal{C}^{K_{\mathrm{s}}}$-smooth regions by g_1^{s} and g_2^{s}. Approximations to the discontinuity $\widehat{b^{\mathrm{s}}}$ and the M-dimensional smooth regions $\widehat{g_1^{\mathrm{s}}}$ and $\widehat{g_2^{\mathrm{s}}}$ may be encoded separately and explicitly. This strategy would have disadvantages for the same reasons mentioned above. In fact, estimating the discontinuity in this scenario would be much harder.

In this paper, we propose representation schemes and algorithms that approximate f^{c} and f^{s} directly in M dimensions. *We emphasize that no explicit knowledge of the functions* b^{c}, b^{s}, g_1^{s}, *or*

g_2^{s} *is required.* We prove that surflet-based approximation techniques and encoding algorithms for f^{c} achieve the optimal decay rates, while our surfprint-based methods for f^{s} achieve the optimal approximation decay rate and a near-optimal rate-distortion decay rate (within a logarithmic factor of the optimal decay rate of Theorem 2). Although we omit the discussion in this paper, our algorithms can be extended to similar piecewise constant and piecewise smooth function spaces. Our spatially localized approach, for example, allows for changes in the variable along which the discontinuity varies (assumed throughout this paper to be x_M as described in Sec. 2.2).

3 The Surflet Dictionary

In this section, we introduce a discrete dictionary of M-dimensional atoms called *surflets* that can be used to construct approximations to a function $f^{\mathrm{c}} \in \mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$. A surflet is a piecewise constant function defined on an M-dimensional dyadic hypercube, where an $(M-1)$-dimensional polynomial specifies the discontinuity. Section 4 describes compression using surflets.

3.1 Motivation — Taylor's theorem

The surflet atoms are motivated by the following property. If d is a function of D variables in $\mathcal{C}^K$ with $K = r + \alpha$, r is a positive integer, and $\alpha \in (0, 1]$, then Taylor's theorem states that

$$\begin{aligned} d(\mathbf{z}+\mathbf{h}) &= d(\mathbf{z}) + \frac{1}{1!}\sum_{i_1=1}^{D} d_{z_{i_1}}(\mathbf{z})h_{i_1} + \frac{1}{2!}\sum_{i_1,i_2=1}^{D} d_{z_{i_1},z_{i_2}}(\mathbf{z})h_{i_1}h_{i_2} + \cdots \\ &\quad + \frac{1}{r!}\sum_{i_1,\ldots,i_r=1}^{D} d_{z_{i_1},\cdots,z_{i_r}}(\mathbf{z})h_{i_1}\cdots h_{i_r} + O(\|\mathbf{h}\|^K), \end{aligned} \tag{1}$$

where $d_{z_1,\cdots,z_\ell}$ refers to the iterated partial derivatives of d with respect to $z_1, \ldots, z_\ell$ in that order. (Note that there are D^ℓ ℓ'th order derivative terms.) Thus, over a small domain, the function d is well approximated using a polynomial of order r (where the polynomial coefficients correspond to the partial derivatives of d evaluated at $\mathbf{z}$).

Clearly, in the case of f^{c}, one method for approximating the discontinuity b^{c} would be to assemble a *piecewise polynomial* approximation, where each polynomial is derived from the local Taylor approximation of b^{c} (let $D = M-1$, $K = K_{\mathrm{d}}$, and $d = b^{\mathrm{c}}$ in the above characterization). These piecewise polynomials can be used to assemble a Horizon-class approximation of the function f^{c}. Surflets provide the M-dimensional framework for constructing such approximations and can be implemented without explicit knowledge of b^{c} or its derivatives.

3.2 Definition

A *dyadic hypercube* $X_j \subseteq [0,1]^M$ at scale $j \in \mathbb{N}$ is a domain that satisfies[5]

$$X_j = [\beta_1 2^{-j}, (\beta_1+1)2^{-j}) \times \cdots \times [\beta_M 2^{-j}, (\beta_M+1)2^{-j})$$

[5]To cover the entire domain $[0,1]^M$, in the case where $(\beta_i+1)2^{-j} = 1$, $i \in \{1,\ldots,M\}$, we replace the half-open interval $[\beta_i 2^{-j}, (\beta_i+1)2^{-j})$ with the closed interval $[\beta_i 2^{-j}, (\beta_i+1)2^{-j}]$.

with $\beta_1, \beta_2, \ldots, \beta_M \in \{0, 1, \ldots, 2^j - 1\}$. We explicitly denote the $(M-1)$-dimensional hypercube *subdomain* of X_j as

$$Y_j = [\beta_1 2^{-j}, (\beta_1 + 1)2^{-j}) \times \cdots \times [\beta_{M-1} 2^{-j}, (\beta_{M-1} + 1)2^{-j}). \tag{2}$$

The *surflet* $s(X_j; p; \cdot)$ is a Horizon-class function over the dyadic hypercube X_j defined through the $(M-1)$-dimensional polynomial p. For $\mathbf{x} \in X_j$ with corresponding $\mathbf{y} = [x_1, x_2, \cdots, x_{M-1}]$,

$$s(X_j; p; \mathbf{x}) = \begin{cases} 1, & x_M \geq p(\mathbf{y}) \\ 0, & \text{otherwise,} \end{cases}$$

where the polynomial $p(\mathbf{y})$ is defined as

$$p(\mathbf{y}) = p_0 + \sum_{i_1=1}^{M-1} p_{1,i_1}\, y_{i_1} + \sum_{i_1,i_2=1}^{M-1} p_{2,i_1,i_2}\, y_{i_1} y_{i_2} + \cdots + \sum_{i_1,\ldots,i_{r_\mathrm{d}}=1}^{M-1} p_{r_\mathrm{d},i_1,i_2,\ldots,i_{r_\mathrm{d}}}\, y_{i_1} y_{i_2} \cdots y_{i_{r_\mathrm{d}}}.$$

We call the polynomial coefficients $\{p_{\ell,i_1,\ldots,i_\ell}\}_{\ell=0}^{r_\mathrm{d}}$ the *surflet coefficients*.[6] We note here that, in some cases, a surflet may be identically 0 or 1 over the entire domain X_j. We sometimes denote a generic surflet by $s(X_j)$, indicating only its region of support.

A surflet $s(X_j)$ approximates the function f^c over the dyadic hypercube X_j. One can cover the entire domain $[0,1]^M$ with a collection of dyadic hypercubes (possibly at different scales) and use surflets to approximate f^c over each of these smaller domains. For $M = 3$, these surflets tiled together look like piecewise polynomial "surfaces" approximating the discontinuity b^c in the function f^c. Figure 2 illustrates a collection of surflets with $M = 2$ and $M = 3$.

3.3 Quantization

We obtain a discrete surflet dictionary $\mathcal{M}(j)$ at scale j by quantizing the set of allowable surflet polynomial coefficients. For $\ell \in \{0, 1, \ldots, r_\mathrm{d}\}$, the surflet coefficient $p_{\ell,i_1,\ldots,i_\ell}$ at scale $j \in \mathbb{N}$ is restricted to values $\{\mu \cdot \Delta_{\ell,j}^{K_\mathrm{d}}\}_{\mu \in \mathbb{Z}}$, where the stepsize satisfies

$$\Delta_{\ell,j}^{K_\mathrm{d}} = 2^{-(K_\mathrm{d} - \ell)j}. \tag{3}$$

The necessary range for μ will depend on the derivative bound Ω (Sec. 2.1). We emphasize that the relevant discrete surflet dictionary $\mathcal{M}(j)$ is *finite* at every scale j.

These quantization stepsizes are carefully chosen to ensure the proper fidelity of surflet approximations without requiring excess bitrate. The key idea is that *higher-order terms can be quantized with lesser precision* without increasing the residual error term in the Taylor approximation (1). In fact, Kolmogorov and Tihomirov [19] implicitly used this concept to establish the metric entropy for bounded uniformly smooth functions.

[6] Because the ordering of the terms $y_{i_1} y_{i_2} \cdots y_{i_\ell}$ in a monomial is irrelevant, only $\binom{\ell+M-2}{\ell}$ monomial coefficients (not $(M-1)^\ell$) need to be encoded for order ℓ. We preserve the slightly redundant notation for ease of comparison with (1).

4 Representation and Coding of Piecewise *Constant* Functions

4.1 Overview

We now propose a surflet-based multiresolution geometric tiling approach to approximate and encode an arbitrary function $f^{\mathrm{c}} \in \mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$. The tiling is arranged on a 2^M-tree, where each node in the tree at scale j corresponds to a hypercube of sidelength 2^{-j}. Each node is labeled with a surflet appropriately chosen from $\mathcal{M}(j)$ and is either a leaf node (hypercube) or has 2^M children nodes (children hypercubes that perfectly tile the volume of the parent hypercube). Leaf nodes provide the actual approximation to the function f^{c}, while interior nodes are useful for predicting and encoding their descendants. This framework enables an *adaptive, multiscale* approximation of f^{c} — many small surflets can be used at fine scales for complicated regions, while few large surflets will suffice to encode simple regions of f^{c} (such as those containing all 0 or 1). Figure 3 shows surflet tiling approximations for $M = 2$ and $M = 3$.

Section 4.2 discusses techniques for determining the proper surflet at each node. Section 4.3 describes a constructive algorithm for building tree-based surflet approximations. Section 4.4 describes the performance of a simple surflet encoder acting only on the leaf nodes. Section 4.5 presents a more advanced surflet coder, using a top-down predictive technique to exploit the correlation among surflet coefficients. Finally, Sec. 4.6 discusses extensions of our surflet-based representation schemes to broader function classes.

4.2 Surflet selection

Consider a node at scale j that corresponds to a dyadic hypercube X_j, and let Y_j be the $(M-1)$-dimensional subdomain of X_j as defined in (2).

We first examine a situation where the coder is provided with explicit information about the discontinuity b^{c} and its derivatives. In this case, determination of the surflet at the node that corresponds to X_j can proceed as implied by Sec. 3. The coder constructs the Taylor expansion of b^{c} around any point $\mathbf{y} \in Y_j$ and quantizes the polynomial coefficients (3). We choose

$$\mathbf{y}_{ep} = \left[\left(\beta_1 + \frac{1}{2}\right) 2^{-j}, \left(\beta_2 + \frac{1}{2}\right) 2^{-j}, \ldots, \left(\beta_{M-1} + \frac{1}{2}\right) 2^{-j} \right]$$

and call this an *expansion point*. We refer to the resulting surflet as the *quantized Taylor surflet*. From (1), it follows that the squared-L_2 error between f^{c} and the quantized Taylor surflet approximation $s(X_j)$ (which equals the X_j-clipped L_1 error between b^{c} and the polynomial defining $s(X_j)$) obeys

$$\|f^{\mathrm{c}} - s(X_j)\|^2_{L_2(X_j)} = \int_{X_j} (f^{\mathrm{c}} - s(X_j))^2 = O\left(2^{-j(K_{\mathrm{d}}+M-1)}\right). \tag{4}$$

However, as discussed in Sec. 2.4, our coder is not provided with explicit information about b^{c}. Therefore, approximating functions in $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ using Taylor surflets is impractical.[7]

We now define a technique for obtaining a surflet estimate directly from the function f^{c}. We assume that there exists a method to compute the squared-L_2 error $\|f^{\mathrm{c}} - s(X_j)\|^2_{L_2(X_j)}$ between

[7] We refer the reader to our technical report [27] for a thorough treatment of Taylor surflet-based approximation of piecewise constant multi-dimensional functions.

a given surflet $s(X_j)$ and the function f^c on the dyadic block X_j. In such a case, we can search the finite surflet dictionary $\mathcal{M}(j)$ for the minimizer of this error *without explicit knowledge of* b^c. We refer to the resulting surflet as the *native* L_2*-best surflet*. This surflet will necessarily obey (4) as well. Section 4.4 discusses the coding implications of using L_2-best surflets from $\mathcal{M}(j)$. Using native L_2-best surflets over dyadic blocks X_j achieves near-optimal performance.

As will be made apparent in Sec. 4.5, in order to achieve optimal performance, a coder must exploit correlations among nearby surflets. Unfortunately, these correlations may be difficult to exploit using native L_2-best surflets. The problem arises because surflets with small X_j-active regions (Sec. 2.3) may be close in L_2 distance over X_j yet have vastly different underlying polynomial coefficients. (These coefficients are used explicitly in our encoding strategy.)

To resolve this problem, we compute L_2-best surflet fits to f^c over the *L-extension* of each dyadic hypercube X_j. That is, if $X_j = [\beta_1 2^{-j}, (\beta_1+1)2^{-j}) \times \cdots \times [\beta_M 2^{-j}, (\beta_M+1)2^{-j})$ then the L-extension of X_j is defined to be

$$X_j^L = [(\beta_1 - L)2^{-j}, (\beta_1 + 1 + L)2^{-j}) \times \cdots \times [(\beta_M - L)2^{-j}, (\beta_M + 1 + L)2^{-j}),$$

where $L > 0$ is an extension factor (designed to expand the domain of analysis and increase correlations between scales).[8] An *L-extended surflet* is a surflet from $\mathcal{M}(j)$ that is now defined over X_j^L whose polynomial discontinuity has a non-empty X_j-active region. We define the *L-extended surflet dictionary* $\mathcal{R}_L(j)$ to be the set of L-extended surflets from $\mathcal{M}(j)$ plus the all-zero and all-one surflets $s(X_j) = 0$ and $s(X_j) = 1$. An *L-extended* L_2*-best surflet* fit to f^c over X_j is then defined to be the L_2-best surflet to f^c *over* X_j^L chosen from $\mathcal{R}_L(j)$. Note that even though extended surflets are defined over extended domains X_j^L, they are used to approximate the function only over the associated native domains X_j. Such extended surflet fits (over extended domains) provide sufficient mathematical constraints for a coder to relate nearby surflets, since extended surflets that are close in terms of squared-L_2 distance over X_j^L have similar polynomial coefficients (even if extended surflets have small X_j-active regions, they have large X_j^L-active regions). In Sec. 4.5, we describe a coder that uses extended surflets from $\mathcal{R}_L(j)$ to achieve optimal performance.

4.3 Tree-based surflet approximations

The surflet dictionary consists of M-dimensional atoms at various scales. Thus, a 2^M-tree offers a natural topology for arranging the surflets used in an approximation. Specifically, each node at scale j in a 2^M-tree is labeled by a surflet that approximates the corresponding dyadic hypercube region X_j of the function f^c. This surflet can be assigned according to any of the procedures outlined in Sec. 4.2.

Given a method for assigning a surflet to each tree node, it is also necessary to determine the proper dyadic segmentation for the tree approximation. This can be accomplished using the CART algorithm, which is based on dynamic programming, in a process known as *tree-pruning* [8, 28]. Tree-pruning proceeds from the bottom up, determining whether to prune the tree beneath each node (causing it to become a leaf node). Various criteria exist for making such a decision. In particular, the approximation-theoretic optimal segmentation can be obtained by minimizing the Lagrangian cost $D+\lambda N$ for a penalty term λ. Similarly, the Lagrangian rate-distortion cost $D+\lambda R$ can be used to obtain the optimal rate-distortion segmentation.

[8] If necessary, each L-extension is truncated to the hypercube $[0,1]^M$.

We summarize the construction of a surflet-based approximation as follows:

Surflet-based approximation

- **Choose scale:** Choose a maximal scale $J \in \mathbb{Z}$ for the 2^M-tree.
- **Label all nodes:** For each scale $j = 0, 1, \ldots, J$, label all nodes at scale j with either a native or an extended L_2-best surflet chosen appropriately from either discrete dictionary of surflets $\mathcal{M}(j)$ or $\mathcal{R}_L(j)$.
- **Prune tree:** Starting at the second-finest scale $j = J - 1$, determine whether each node at scale j should be pruned (according to an appropriate pruning rule). Then proceed up to the root of the tree, i.e., until $j = 0$.

The approximation performance of this algorithm is described in the following theorem, which is proved in Appendix C.

Theorem 3 *Using either quantized Taylor surflets or L_2-best surflets (extended or native), a surflet tree-pruned approximation of an element $f^{\mathrm{c}} \in \mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ achieves the optimal asymptotic approximation rate of Theorem 1:*

$$\left\| f^{\mathrm{c}} - \widehat{f}^{\mathrm{c}}_N \right\|^2_{L_2} \lesssim \left(\frac{1}{N} \right)^{\frac{K_{\mathrm{d}}}{M-1}}.$$

We discuss computational issues associated with finding best-fit surflets in Sec. 6.5, where we also present results from simulations.

4.4 Leaf encoding

An initial approach toward surflet encoding would involve specification of the tree segmentation map (which denotes the location of the leaf nodes) along with the quantized surflet coefficients at each leaf node. The rate-distortion analysis in Appendix D then yields the following result.

Theorem 4 *Using either quantized Taylor surflets or L_2-best surflets (extended or native), a surflet leaf-encoder applied to an element $f^{\mathrm{c}} \in \mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ achieves the following rate-distortion performance*

$$\left\| f^{\mathrm{c}} - \widehat{f}^{\mathrm{c}}_R \right\|^2_{L_2} \lesssim \left(\frac{\log R}{R} \right)^{\frac{K_{\mathrm{d}}}{M-1}}.$$

Comparing with Theorem 1, this simple coder is *near-optimal* in terms of rate-distortion performance. The logarithmic factor is due to the fact that it requires $O(j)$ bits to encode each surflet at scale j. In Sec. 4.5, we propose an alternative coder that requires only a constant number of bits to encode each surflet.

4.5 Top-down predictive encoding

Achieving the optimal performance of Theorem 1 requires a more sophisticated coder that can exploit the correlation among nearby surflets. We now briefly describe a top-down surflet coder that predicts surflet parameters from previously encoded values.

Top-down predictive surflet coder

- **Encode root node:** Encode the best surflet fit $s([0,1]^M)$ to the hypercube $[0,1]^M$. Encode a flag (1-bit) specifying whether this node is interior or a leaf. Set $j \leftarrow 0$.
- **Predict surflets from parent scale:** For every interior node/hypercube X_j at scale j, partition its domain into 2^M children hypercubes at scale $j+1$. Compute the polynomial coefficients on each child hypercube X_{j+1} that agree with the encoded parent surflet $s(X_j^L)$. These serve as "predictions" for the polynomial coefficients at the child.
- **Encode innovations at child nodes:** For each predicted polynomial coefficient, encode the discrepancy with the L-extended surflet fit $s(X_{j+1}^L)$.
- **Descend tree:** Set $j \leftarrow j+1$ and repeat until no interior nodes remain.

This top-down predictive coder encodes an entire tree segmentation starting with the root node, and proceeding from the top down. Given an L-extended surflet $s(X_j^L)$ at an interior node at scale j, we show in Appendix E that the number of possible L-extended surflets from $\mathcal{R}_L(j)$ that can be used for approximation at scale $j+1$ is *constant*, independent of the scale j. Thus, given a best-fit surflet at scale 0, a constant number of bits is required to encode each surflet at subsequent scales. This prediction is possible because L-extended surflets are defined over L-extended domains, which ensures coherency between the surflet fits (and polynomial coefficients) at a parent and child node.

We note that predicting L-extended best-fit surflets to dyadic hypercube regions around the borders of $[0,1]^M$ may not be possible with a constant number of bits when the discontinuity is not completely contained within the dyadic hypercube. However, we make the mild simplifying assumption that the intersections of the discontinuity with the hyperplanes $x_M = 0$ or $x_M = 1$ can be contained within $O(2^{(M-2)j})$ hypercubes at each scale j. Therefore, using $O(K_{\mathrm{d}} j)$ bits to encode such "border" dyadic hypercubes (with the discontinuity intersecting $x_M = 0$ or $x_M = 1$) does not affect the asymptotic rate-distortion performance of the top-down predictive coder. In Appendix E, we prove the following theorem.

Theorem 5 *The top-down predictive coder applied to an element $f^{\mathrm{c}} \in \mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ using L-extended L_2-best surflets from $\mathcal{R}_L(j)$ achieves the optimal rate-distortion performance of Theorem 1:*

$$\left\| f^{\mathrm{c}} - \widehat{f_R^{\mathrm{c}}} \right\|_{L_2}^2 \lesssim \left(\frac{1}{R} \right)^{\frac{K_{\mathrm{d}}}{M-1}}.$$

Although only the leaf nodes provide the ultimate approximation to the function, the additional information encoded at interior nodes provides the key to efficiently encoding the leaf nodes. In addition, unlike the surflet leaf-encoder of Sec. 4.3, this top-down approach yields a *progressive* bitstream — the early bits encode a low-resolution (coarse scale) approximation, which is then refined using subsequent bits.

4.6 Extensions to broader function classes

Our results for classes of functions that contain a single discontinuity can be extended to spaces of signals that contain multiple discontinuities. Functions containing multiple discontinuities that do not intersect can be represented using the surflet-based approximation scheme described in Sec. 4.3 at the optimal asymptotic approximation rate. This is because at a sufficiently high scale, dyadic hypercubes that tile signals containing multiple non-intersecting discontinuities contain at most one discontinuity.

Analysis of the surflet-based approximation scheme of Sec. 4.3 applied to signals containing intersecting discontinuities is more involved. Let $f_\sharp^{\mathrm{c}}$ be an M-dimensional piecewise constant function containing two $(M-1)$-dimensional $\mathcal{C}^{K_{\mathrm{d}}}$-smooth discontinuities that intersect each other (the analysis that follows can easily be extended to allow for more than two intersecting discontinuities). Note that the intersection of $(M-1)$-dimensional functions forms an $(M-2)$-dimensional manifold. Again, we make the mild simplifying assumption that the intersection of the discontinuities can be contained in $O(2^{(M-2)j})$ hypercubes at each scale j. The following theorem describes the approximation performance achieved by the scheme in Sec. 4.3 applied to $f_\sharp^{\mathrm{c}}$. A consequence of this theorem is that there exists a smoothness threshold $K_{\mathrm{d}}^{\mathrm{th}}$ that defines the boundary between optimal and sub-optimal approximation performance.

Theorem 6 *Using either quantized Taylor surflets or L_2-best surflets (extended or native), the approximation scheme of Sec. 4.3 applied to a piecewise constant M-dimensional function $f_\sharp^{\mathrm{c}}$ that contains two intersecting $\mathcal{C}^{K_{\mathrm{d}}}$-smooth $(M-1)$-dimensional discontinuities achieves performance given by:*

- $M > 2, K_{\mathrm{d}} \leq \frac{2(M-1)}{M-2}$*:*
$$\left\| f_\sharp^{\mathrm{c}} - \widehat{f_{\sharp,N}^{\mathrm{c}}} \right\|_{L_2}^2 \lesssim \left(\frac{1}{N}\right)^{\frac{K_{\mathrm{d}}}{M-1}}.$$
- $M > 2, K_{\mathrm{d}} > \frac{2(M-1)}{M-2}$*:*
$$\left\| f_\sharp^{\mathrm{c}} - \widehat{f_{\sharp,N}^{\mathrm{c}}} \right\|_{L_2}^2 \lesssim \left(\frac{1}{N}\right)^{\frac{2}{M-2}}.$$
- $M = 2$, *any* K_{d}*:*
$$\left\| f_\sharp^{\mathrm{c}} - \widehat{f_{\sharp,N}^{\mathrm{c}}} \right\|_{L_2}^2 \lesssim \left(\frac{1}{N}\right)^{\frac{K_{\mathrm{d}}}{M-1}}.$$

Thus, the representation scheme in Sec. 4.3 achieves optimal approximation performance for $M = 2$ even in the presence of intersecting discontinuities, while it achieves optimal performance for $M > 2$ up to a smoothness threshold of $K_{\mathrm{d}}^{\mathrm{th}} = \frac{2(M-1)}{M-2}$ (for $K_{\mathrm{d}} > K_{\mathrm{d}}^{\mathrm{th}}$, the scheme performs sub-optimally: $\|f_\sharp^{\mathrm{c}} - \widehat{f_{\sharp,N}^{\mathrm{c}}}\|_{L_2}^2 \lesssim \left(\frac{1}{N}\right)^{\frac{K_{\mathrm{d}}^{\mathrm{th}}}{M-1}}$). This performance of the approximation scheme for $M > 2$ is still superior to that of wavelets, which have $K_{\mathrm{d}}^{\mathrm{th,wl}} = 1$. The reason for this difference in performance between the cases $M = 2$ and $M > 2$ is that intersections of discontinuities when $M =$

2 correspond to points,[9] while intersections in higher dimensions correspond to low-dimensional manifolds. Hence, the number of hypercubes that contain intersections in the two-dimensional case is constant with scale, whereas the number of hypercubes that contain the intersections when $M > 2$ grows exponentially with scale. The analysis above can clearly be extended to prove analogous results for functions containing piecewise $\mathcal{C}^{K_\mathrm{d}}$-smooth discontinuities.

Future work will focus on improving the threshold $K_\mathrm{d}^{\mathrm{th}}$ for the case $M > 2$. In order to achieve optimal performance for $M > 2$, one may need a dictionary containing regular surflets and specially-designed "intersection" surflets that are specifically tailored for intersections. In addition to classes of functions containing multiple intersecting discontinuities, our representation scheme may also be adapted to function spaces where the transient direction of the discontinuity is not fixed to be x_M. This is possible due to the localized nature of our surflet-based approximations.

5 Representation and Coding of Piecewise *Smooth* Functions

In this section, we extend our coding strategies for piecewise constant functions to encoding an arbitrary element f^s from the class $\mathcal{F}_\mathrm{S}(M, K_\mathrm{d}, K_\mathrm{s})$ of piecewise smooth functions.

5.1 Motivation

For a $\mathcal{C}^{K_\mathrm{s}}$-smooth function f in M dimensions, a wavelet basis with sufficient vanishing moments [22] provides approximations at the optimal rate — $\|f - \widehat{f}_N\|_{L_2}^2 \lesssim \left(\frac{1}{N}\right)^{\frac{2K_\mathrm{s}}{M}}$. Even if one introduces a finite number of point singularities into the M-dimensional $\mathcal{C}^{K_\mathrm{s}}$-smooth function, wavelet-based approximation schemes still attain the optimal rate. Wavelets succeed in approximating smooth functions because most of the wavelet coefficients have small magnitudes and can thus be neglected. Moreover, an arrangement of wavelet coefficients on the nodes of a tree leads to an interesting consequence: wavelet coefficients used in the approximation of M-dimensional smooth functions are *coherent* — often, if a wavelet coefficient has small magnitude, then its children coefficients also have small magnitude. These properties of the wavelet basis have been exploited in state-of-the-art wavelet-based image coders [4, 5].

Although wavelets approximate smooth functions well, the wavelet basis is not well-equipped to approximate functions containing higher-dimensional manifold discontinuities. Wavelets also do not take advantage of any structure (such as smoothness) that the $(M-1)$-dimensional discontinuity might have, and therefore many high-magnitude coefficients are often required to represent discontinuities [16]. Regardless of the smoothness order of the discontinuity, the approximation rate achieved by wavelets remains the same.

Despite this drawback, we desire a wavelet domain solution to approximate $f^\mathrm{s} \in \mathcal{F}_\mathrm{S}(M, K_\mathrm{d}, K_\mathrm{s})$ because most of the function f^s is smooth in M dimensions, except for an $(M-1)$-dimensional discontinuity. In order to solve the problem posed by the discontinuity, we propose the addition of *surfprint* atoms to the dictionary of wavelet atoms. A surfprint is a weighted sum of wavelet basis functions derived from the projection of a piecewise polynomial surflet atom (an $(M-1)$-dimensional polynomial discontinuity separating two M-dimensional polynomial regions) onto a subspace in the wavelet domain (see Fig. 5 for an example in 2D). Surfprints possess all the properties that make

[9] Our analysis also applies to "T-junctions" in images, where one edge terminates at its intersection with another.

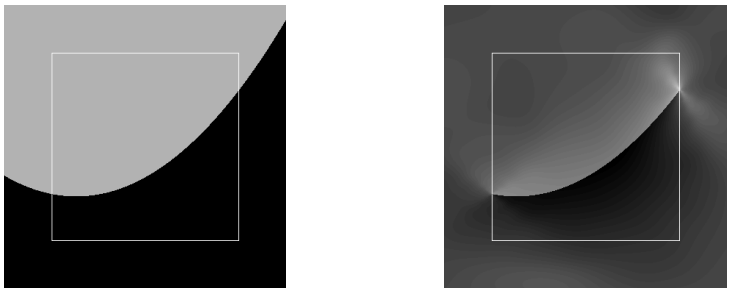


Figure 5: *Example surflet and the corresponding surfprint. The white box is the dyadic hypercube in which we define the surflet; note that the removal of coarse scale and neighboring wavelets causes the surfprint to appear different from the surflet.*

surflets well-suited to represent discontinuities. In addition, surfprints coherently model wavelet coefficients that correspond to discontinuities. Thus, we obtain a single unified wavelet-domain framework that is well-equipped to sparsely represent both discontinuities and smooth regions.

The rest of this section is devoted to the definition of surfprints and their use in a wavelet domain framework to represent and encode approximations to elements of $\mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$. We do not discuss the extension of our results to classes of piecewise smooth signals containing multiple intersecting discontinuities, but note that such an analysis would be similar to that described in Sec. 4.6.

5.2 Surfprints

Let X_{J_o} be a dyadic hypercube at scale J_o. Let v_1, v_2 be M-dimensional polynomials of degree $r_{\mathrm{s}}^{\mathrm{sp}}$, and let v be an M-dimensional function as follows:

$$v_1, v_2, v : X_{J_o} \to \mathbb{R}.$$

Let q be an $(M-1)$-dimensional polynomial of degree $r_{\mathrm{d}}^{\mathrm{sp}}$:

$$q : Y_{J_o} \to \mathbb{R}.$$

As defined in Sec. 2.2, let $\mathbf{x} \in X_{J_o}$ and let $\mathbf{y}$ denote the first $M-1$ elements of $\mathbf{x}$. Let the M-dimensional piecewise polynomial function v be defined as follows:

$$v(\mathbf{x}) = \begin{cases} v_1(\mathbf{x}), & x_M \geq q(\mathbf{y}) \\ v_2(\mathbf{x}), & x_M < q(\mathbf{y}). \end{cases}$$

Next, we describe how this piecewise polynomial function is projected onto a wavelet subspace to obtain a surfprint atom. Let $\mathcal{W}$ be a compactly supported wavelet basis in M dimensions with $K_{\mathrm{s}}^{\mathrm{wl}}$ vanishing moments. A surfprint $\mathrm{sp}(v, X_{J_o}, \mathcal{W})$ is a weighted sum of wavelet basis functions with the weights derived by projecting the piecewise polynomial v onto the subtree of basis functions whose idealized supports nest in the hypercube X_{J_o}:

$$\mathrm{sp}(v, X_{J_o}, \mathcal{W}) = \sum_{j \geq J_o, X_j \subseteq X_{J_o}} \left\langle v, w_{X_j} \right\rangle w_{X_j}, \tag{5}$$

where w_{X_j} represents the wavelet basis function having idealized compact support on the hypercube X_j. (The actual support of w_{X_j} may extend slightly beyond X_j.) The hypercube X_{J_o} thus defines the root node (or coarsest scale) of the surfprint atom.

We propose an approximation scheme in Sec. 5.5 where we use wavelet atoms to represent uniformly smooth regions of f^{s} and surfprint atoms to represent regions through which the discontinuity passes. Before presenting our approximation scheme, we begin in Sec. 5.3 by describing how to choose the surfprint polynomial degrees $r_{\mathrm{s}}^{\mathrm{sp}}$ and $r_{\mathrm{d}}^{\mathrm{sp}}$ and the number of vanishing moments $K_{\mathrm{s}}^{\mathrm{wl}}$ for the wavelet basis.

5.3 Vanishing moments and polynomial degrees

In general, due to Taylor's theorem, when approximating elements $f^{\mathrm{s}} \in \mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$, the required surfprint polynomial degrees and wavelet vanishing moments are determined by the orders of smoothness K_{d} and K_{s}:

$$K_{\mathrm{s}}^{\mathrm{wl}} \geq K_{\mathrm{s}}, \quad r_{\mathrm{d}}^{\mathrm{sp}} = \lceil K_{\mathrm{d}} - 1 \rceil, \quad \text{and } r_{\mathrm{s}}^{\mathrm{sp}} = \lceil K_{\mathrm{s}} - 1 \rceil.$$

However, the exponent in the expression of Theorem 2 for the optimal approximation rate for $\mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$ indicates that for every $(K_{\mathrm{d}}, K_{\mathrm{s}})$, either the $(M-1)$-dimensional discontinuity or the M-dimensional smooth region dominates the decay rate. For instance, in two dimensions, the smaller of the two smoothness orders K_{d} and K_{s} defines the decay rate.[10] This implies that the surfprint polynomial degrees and/or the number of wavelet vanishing moments can be relaxed (as if either the discontinuity or the smooth regions had a lower smoothness order), without affecting the approximation rate.

Rather than match the surfprint parameters directly to the smoothness orders K_{d} and K_{s}, we let $K_{\mathrm{d}}^{\mathrm{sp}}$ and $K_{\mathrm{s}}^{\mathrm{sp}}$ denote the *operational smoothness orders* to which the surfprint parameters are matched. These operational smoothness orders are selected to ensure the best approximation or rate-distortion performance. The detailed derivations of Appendix G and Appendix H yield the following values for the operational smoothness orders:

- **Discontinuity dominates:** In this case, $\frac{K_{\mathrm{d}}}{M-1} < \frac{2K_{\mathrm{s}}}{M}$. We let $K_{\mathrm{d}}^{\mathrm{sp}} = K_{\mathrm{d}}$ and choose $K_{\mathrm{s}}^{\mathrm{sp}} \in [\frac{K_{\mathrm{d}}-1}{2}, K_{\mathrm{s}}]$ and $K_{\mathrm{s}}^{\mathrm{wl}} \in [\frac{K_{\mathrm{d}}M}{2(M-1)}, K_{\mathrm{s}}]$.
- **Smooth regions dominate:** In this case, $\frac{2K_{\mathrm{s}}}{M} < \frac{K_{\mathrm{d}}}{M-1}$. We let $K_{\mathrm{s}}^{\mathrm{wl}} = K_{\mathrm{s}}$, and choose $K_{\mathrm{s}}^{\mathrm{sp}} \in [K_{\mathrm{s}}(1 - \frac{1}{M}) - \frac{1}{2}, K_{\mathrm{s}}]$ and $K_{\mathrm{d}}^{\mathrm{sp}} \in [\frac{2K_{\mathrm{s}}(M-1)}{M}, K_{\mathrm{d}}]$.
- **Both contribute equally:** In this case, $\frac{2K_{\mathrm{s}}}{M} = \frac{K_{\mathrm{d}}}{M-1}$. We let $K_{\mathrm{s}}^{\mathrm{wl}} = K_{\mathrm{s}}$, $K_{\mathrm{d}}^{\mathrm{sp}} = K_{\mathrm{d}}$, and choose $K_{\mathrm{s}}^{\mathrm{sp}} \in [K_{\mathrm{s}}(1 - \frac{1}{M}) - \frac{1}{2}, K_{\mathrm{s}}]$.

The surfprint polynomial degrees are given by

$$r_{\mathrm{d}}^{\mathrm{sp}} = \lceil K_{\mathrm{d}}^{\mathrm{sp}} - 1 \rceil \quad \text{and} \quad r_{\mathrm{s}}^{\mathrm{sp}} = \lceil K_{\mathrm{s}}^{\mathrm{sp}} - 1 \rceil.$$

[10] We note also that in the case where the functions g_1 and g_2, which characterize f^{s} above and below the discontinuity, have differing orders of smoothness, the smaller smoothness order will determine both the achievable approximation rates and the appropriate approximation strategies.

Therefore, if $\lceil K_{\mathrm{d}}^{\mathrm{sp}} - 1 \rceil < \lceil K_{\mathrm{d}} - 1 \rceil$ and $\lceil K_{\mathrm{s}}^{\mathrm{sp}} - 1 \rceil < \lceil K_{\mathrm{s}} - 1 \rceil$, then the required surfprint polynomial degrees for optimal approximations are lower than what one would naturally expect. Note that even in the scenario where both terms in the exponent of the approximation rate match, one can choose $K_{\mathrm{s}}^{\mathrm{sp}}$ slightly smaller than K_{s} while still attaining the optimal approximation rate of Theorem 2.

5.4 Quantization

In order to construct a discrete surfprint/wavelet dictionary, we quantize the coefficients of the wavelet and surfprint atoms. The quantization step-size $\Delta^{K_{\mathrm{s}}^{\mathrm{wl}}}$ for the wavelet coefficients depends on the specific parameters of an approximation scheme. We present our prototype approximation scheme and discuss the wavelet coefficient step-sizes in Sec. 5.5 (see (8) below).

The quantization step-size for the surfprint polynomial coefficients of order ℓ at scale j is analogous to the step-size used to construct a discrete surflet dictionary (3):

$$\Delta_{\ell,j}^{K_{\mathrm{d}}^{\mathrm{sp}}} = 2^{-(K_{\mathrm{d}}^{\mathrm{sp}} - \ell)j} \tag{6}$$

and

$$\Delta_{\ell,j}^{K_{\mathrm{s}}^{\mathrm{sp}}} = 2^{-(K_{\mathrm{s}}^{\mathrm{sp}} - \ell)j}. \tag{7}$$

As before, the key idea is that higher-order polynomial coefficients can be quantized with lesser precision without affecting the error term in the Taylor approximation (1).

5.5 Surfprint-based approximation

We present a tree-based representation scheme using quantized wavelet and surfprint atoms and prove that this scheme achieves the optimal approximation rate for every function $f^{\mathrm{s}} \in \mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$. Let $\mathcal{W}$ be a compactly supported wavelet basis in M dimensions with $K_{\mathrm{s}}^{\mathrm{wl}}$ vanishing moments, as defined in Sec. 5.3. Consider the decomposition of f^{s} into the wavelet basis vectors: $f^{\mathrm{s}} = \sum_j \langle f^{\mathrm{s}}, w_{X_j} \rangle w_{X_j}$. The wavelet coefficients $\langle f^{\mathrm{s}}, w_{X_j} \rangle$ are quantized according to the step-size $\Delta^{K_{\mathrm{s}}^{\mathrm{wl}}}$ defined below. Let these wavelet atoms be arranged on the nodes of a 2^M-tree. We classify the nodes based on the idealized support of the corresponding wavelet basis functions. Nodes whose supports X_j are intersected by the discontinuity b^{s} are called Type D nodes. All other nodes (over which f^{s} is smooth) are classified as Type S. Consider now the following surfprint approximation strategy:[11]

Surfprint approximation

- **Choose scales and wavelet quantization step-size:** Choose a maximal scale $J \in \mathbb{Z}$ and $m, n \in \mathbb{Z}$ such that $\frac{m}{n} = \frac{M}{M-1}$ and both m and n divide J. The quantization step-size for wavelet coefficients at all scales j is given by:

$$\Delta^{K_{\mathrm{s}}^{\mathrm{wl}}} = 2^{-\frac{J}{m}(K_{\mathrm{s}}^{\mathrm{wl}} + \frac{M}{2})} \tag{8}$$

  and thus depends only on the maximal scale J and the parameter m.

[11] The wavelet decomposition actually has $2^M - 1$ distinct directional subbands; we assume here that each is treated identically. Also we assume the scaling coefficient at the coarsest scale $j = 0$ is encoded as side information with negligible cost.

- **Prune tree:** Keep all wavelet nodes up to scale $\frac{J}{m}$; from scale $\frac{J}{m}$ to scale $\frac{J}{n}$, prune the tree at all Type S nodes (discarding those wavelet coefficients and their descendant subtrees).
- **Select surfprint atoms:** At scale $\frac{J}{n}$ replace the wavelet atom at each Type D discontinuity node and its descendant subtree (up to depth J) by a quantized surfprint atom chosen appropriately from the dictionary with $J_o = \frac{J}{n}$ in (5):
  - M-dimensional polynomials: Choose M-dimensional polynomials v_1 and v_2 of degree $r_{\mathrm{s}}^{\mathrm{sp}} = \lceil K_{\mathrm{s}}^{\mathrm{sp}} - 1 \rceil$. These polynomials should approximate the M-dimensional smooth regions up to an absolute (pointwise) error of $O\left(2^{\frac{-K_{\mathrm{s}}^{\mathrm{sp}} J}{n}}\right)$. The existence of such polynomials is guaranteed by Taylor's theorem (1) (let $D = M$, $K = K_{\mathrm{s}}^{\mathrm{sp}}$, and $r = r_{\mathrm{s}}^{\mathrm{sp}}$) and the quantization scheme (7).
  - $(M-1)$-dimensional polynomial: Choose an $(M-1)$-dimensional polynomial q of degree $r_{\mathrm{d}}^{\mathrm{sp}} = \lceil K_{\mathrm{d}}^{\mathrm{sp}} - 1 \rceil$ such that the discontinuity is approximated up to an absolute error of $O\left(2^{\frac{-K_{\mathrm{d}}^{\mathrm{sp}} J}{n}}\right)$. The existence of such a polynomial is guaranteed by Taylor's theorem (1) (let $D = M-1$, $K = K_{\mathrm{d}}^{\mathrm{sp}}$, and $r = r_{\mathrm{d}}^{\mathrm{sp}}$) and the quantization scheme of (6).

The following theorem summarizes the performance analysis for such surfprint approximations (see Appendix G for the proof).

Theorem 7 *A surfprint-based approximation of an element $f^{\mathrm{s}} \in \mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$ as presented above achieves the optimal asymptotic approximation rate of Theorem 2:*

$$\left\| f^{\mathrm{s}} - \widehat{f_N^{\mathrm{s}}} \right\|_{L_2}^2 \lesssim \left(\frac{1}{N}\right)^{\min\left(\frac{K_{\mathrm{d}}}{M-1}, \frac{2K_{\mathrm{s}}}{M}\right)}.$$

An approximation scheme that uses the *best* configuration of N wavelet and surfprint atoms in the L_2 sense would perform at least as well as the scheme suggested above. Hence, surfprint approximation algorithms designed to choose the best N-term approximation (even without explicit knowledge of the discontinuity or the M-dimensional smooth regions) will achieve the optimal approximation rate of Theorem 2.

5.6 Encoding a surfprint/wavelet approximation

We now consider the problem of encoding the tree-based approximation of Sec. 5.5. A simple top-down coding scheme that specifies the pruned tree topology, quantized wavelet coefficients, and surfprint parameters achieves a near-optimal rate-distortion performance (see Appendix H for proof).

Theorem 8 *A coding scheme that encodes every element of the surfprint-based approximation of an element $f^{\mathrm{s}} \in \mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$ as presented in Sec. 5.5 achieves the near-optimal asymptotic rate-distortion performance (within a logarithmic factor of the optimal performance of Theorem 2):*

$$\left\| f^{\mathrm{s}} - \widehat{f_R^{\mathrm{s}}} \right\|_{L_2}^2 \lesssim \left(\frac{\log R}{R}\right)^{\min\left(\frac{K_{\mathrm{d}}}{M-1}, \frac{2K_{\mathrm{s}}}{M}\right)}.$$

Repeating the argument of Sec. 5.5, this near optimal rate-distortion performance serves as an upper bound for an encoding scheme that encodes elements of an L_2-best approximation. We will discuss the extension of these theoretical results to the approximation of discrete data and related issues in Sec. 6.3.

6 Extensions to Discrete Data

6.1 Overview

In this section, we consider the problem of representing discrete data obtained by "voxelizing" (pixelizing in 2D dimensions) functions from the classes $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ and $\mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$. Let f be a continuous M-dimensional function. We discretize f according to a vector $\pi = [2^{\pi_1}, \ldots, 2^{\pi_M}] \in \mathbb{Z}^M$, which specifies the number of voxels along each dimension of the discretized M-dimensional function $\widetilde{f_\pi}$. Each entry of $\widetilde{f_\pi}$ is obtained either by averaging f over an M-dimensional voxel or by sampling f at uniformly spaced intervals. (Because of the smoothness characteristics of $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ and $\mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$, both discretization mechanisms provide the same asymptotic performance.) In our analysis, we allow the number of voxels along each dimension to vary in order to provide a framework for analyzing various sampling rates along the different dimensions. Video data, for example, is often sampled differently in the spatial and temporal dimensions. Future research will consider different distortion criteria based on asymmetry in the spatiotemporal response of the human visual system.

For our analysis, we assume that the voxelization vector π is fixed and denote the resulting classes of voxelized functions by $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$ and $\widetilde{\mathcal{F}_{\mathrm{S}}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$. Sections 6.2 and 6.3 describe the sparse representation of elements from $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$ and $\widetilde{\mathcal{F}_{\mathrm{S}}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$, respectively. In Sec. 6.4, we discuss the impact of discretization effects on fine scale approximations. Finally, we present our simulation results in Sec. 6.5.

6.2 Representing and encoding elements of $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$

Suppose $f^{\mathrm{c}} \in \mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ and let $\widetilde{f_\pi^{\mathrm{c}}} \in \widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$ be its discretization. (We view $\widetilde{f_\pi^{\mathrm{c}}}$ as a function on the continuous domain $[0,1]^M$ that is constant over each voxel.) The process of voxelization affects the ability to approximate elements of $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$. At coarse scales, however, much of the intuition for coding $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ can be retained. In particular, we can bound the distance from $\widetilde{f_\pi^{\mathrm{c}}}$ to f^{c}. We note that $\widetilde{f_\pi^{\mathrm{c}}}$ differs from f^{c} only over voxels through which b passes. Because each voxel has size $2^{-\pi_1} \times 2^{-\pi_2} \cdots \times 2^{-\pi_M}$, the number of voxels intersected by b is $O\left(2^{\sum_{i=1}^{M-1} \pi_i} \left\lceil \left(\Omega \cdot 2^{-\min(\pi_i)_{i=1}^{M-1}}\right) / \left(2^{-\pi_M}\right) \right\rceil\right)$, where Ω is the universal derivative bound (Sec. 2.1). The squared-L_2 distortion incurred on each such voxel (assuming only that the voxelization process is bounded and local) is $O(2^{-(\pi_1 + \cdots + \pi_M)})$. Summing over all voxels it follows that the (nonsquared) L_2 distance obeys

$$\left\| f^{\mathrm{c}} - \widetilde{f_\pi^{\mathrm{c}}} \right\|_{L_2([0,1]^M)} < C_1 \cdot 2^{-(\min \pi_i)/2} \tag{9}$$

where the minimum is taken over all $i \in \{1, \ldots, M\}$.

Now we consider the problem of encoding elements of $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$. At a particular bitrate R, we know from Theorem 1 that no encoder could represent all elements of $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ using R bits and

incurring L_2 distortion less than $C_2 \cdot \left(\frac{1}{R}\right)^{\frac{K_\mathrm{d}}{2(M-1)}}$. (This lower bound for metric entropy is in effect for R sufficiently large, which we assume to be the case.) Suppose we consider a hypothetical encoder for elements of $\widetilde{\mathcal{F}_\mathrm{C}}(M, K_\mathrm{d})$ that, using R bits, could represent any element with L_2 distortion of $\widetilde{\mathcal{F}_\mathrm{C}}(M, K_\mathrm{d})$ less than some $D_\mathrm{hyp}(R)$. This coder could *also* be used as an encoder for elements of $\mathcal{F}_\mathrm{C}(M, K_\mathrm{d})$ (by voxelizing each function before encoding). This strategy would yield L_2 distortion no worse than $C_1 \cdot 2^{-(\min \pi_i)/2} + D_\mathrm{hyp}(R)$. By applying the metric entropy arguments on $\mathcal{F}_\mathrm{C}(M, K_\mathrm{d})$, we have the following constraint on $D_\mathrm{hyp}(R)$:

$$C_1 \cdot 2^{-(\min \pi_i)/2} + D_\mathrm{hyp}(R) \geq C_2 \cdot \left(\frac{1}{R}\right)^{\frac{K_\mathrm{d}}{2(M-1)}},$$

or equivalently,

$$D_\mathrm{hyp}(R) \geq C_2 \cdot \left(\frac{1}{R}\right)^{\frac{K_\mathrm{d}}{2(M-1)}} - C_1 \cdot 2^{-(\min \pi_i)/2}. \tag{10}$$

This inequality helps establish a rate-distortion bound for the class $\widetilde{\mathcal{F}_\mathrm{C}}(M, K_\mathrm{d})$. At sufficiently low rates, the first term on the RHS dominates, and $\widetilde{\mathcal{F}_\mathrm{C}}(M, K_\mathrm{d})$ faces similar rate-distortion constraints to $\mathcal{F}_\mathrm{C}(M, K_\mathrm{d})$. At high rates, however, the RHS becomes negative, giving little insight into the coding of $\widetilde{\mathcal{F}_\mathrm{C}}(M, K_\mathrm{d})$. This breakdown point occurs when $R \sim 2^{(\min \pi_i)(M-1)/K_\mathrm{d}}$.

We can, in fact, specify a constructive encoding strategy for $\widetilde{\mathcal{F}_\mathrm{C}}(M, K_\mathrm{d})$ that achieves the optimal compression rate up to this breakdown point. We construct a dictionary of discrete surflet atoms by voxelizing the elements of the continuous quantized surflet dictionary. Assuming there exists a technique to find discrete ℓ_2-best surflet fits to $\widetilde{f_\pi^\mathrm{c}}$, the tree-based algorithm described in Sec. 4.3 can simply be used to construct an approximation $\widehat{\widetilde{f_\pi^\mathrm{c}}}$.

Theorem 9 *While $R \lesssim 2^{(\min \pi_i)(M-1)/K_\mathrm{d}}$, the top-down predictive surflet coder from Sec. 4.5 applied to encode the approximation $\widehat{\widetilde{f_\pi^\mathrm{c}}}$ to $\widetilde{f_\pi^\mathrm{c}}$ using discrete ℓ_2-best surflets achieves the rate-distortion performance*

$$\left\| \widetilde{f_\pi^\mathrm{c}} - \widehat{\widetilde{f_\pi^\mathrm{c}}} \right\|_{L_2}^2 \lesssim \left(\frac{1}{R}\right)^{\frac{K_\mathrm{d}}{M-1}}.$$

As detailed in the proof of this theorem (see Appendix I), the breakdown point occurs when using surflets at a critical scale $J_\mathrm{vox} = \frac{\min \pi_i}{K_\mathrm{d}}$. Up to this scale, all of the familiar approximation and compression rates hold. Beyond this scale, however, voxelization effects dominate. An interesting corollary to Theorem 9 is that, due to the similarities up to scale J_vox, the discrete approximation $\widehat{\widetilde{f_\pi^\mathrm{c}}}$ itself provides an effective approximation to the function f^c.

Corollary 10 *While $R \lesssim 2^{(\min \pi_i)(M-1)/K_\mathrm{d}}$, the discrete approximation $\widehat{\widetilde{f_\pi^\mathrm{c}}}$ provides an approximation to f^c with the following rate-distortion performance:*

$$\left\| f^\mathrm{c} - \widehat{\widetilde{f_\pi^\mathrm{c}}} \right\|_{L_2}^2 \lesssim \left(\frac{1}{R}\right)^{\frac{K_\mathrm{d}}{M-1}}.$$

The details of the proof appear in Appendix J. While we have provided an effective strategy for encoding elements of $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$ at sufficiently low rates (using surflets at scales $j \leq J_{\mathrm{vox}}$), this leaves open the question of how to code $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$ at higher rates. Unfortunately, (10) does not offer much insight. In particular, it is not clear whether surflets are an efficient strategy for encoding $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$ beyond scale J_{vox}. We revisit this issue in Sec. 6.4.

6.3 Representing and encoding elements of $\widetilde{\mathcal{F}_{\mathrm{S}}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$

Next, let $\widetilde{f_\pi^{\mathrm{s}}}$ be an arbitrary signal belonging to $\widetilde{\mathcal{F}_{\mathrm{S}}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$. Similar arguments apply to the voxelization effects for this class. In order to approximate functions in $\widetilde{\mathcal{F}_{\mathrm{S}}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$, we use a dictionary of compactly supported discrete wavelet basis functions with $K_{\mathrm{s}}^{\mathrm{wl}}$ vanishing moments and discrete surfprint atoms. A discrete surfprint atom is derived by projecting a discrete piecewise polynomial surflet atom onto a subspace of the discrete wavelet basis.

We use the scheme described in Sec. 5.5 with $\frac{J_{\mathrm{vox}}}{n} = \frac{\min(\pi_i)}{\min\left(K_{\mathrm{d}}^{\mathrm{sp}}, 2K_{\mathrm{s}}^{\mathrm{sp}}+1\right)}$ to approximate $\widetilde{f_\pi^{\mathrm{s}}}$ by $\widehat{\widetilde{f_\pi^{\mathrm{s}}}}$. Using (40), (41), and (42), this scale corresponds to a range of bitrates up to $O(J_{\mathrm{vox}} 2^{(M-1)\frac{J_{\mathrm{vox}}}{n}})$. Within this range, the approximation is encoded as described in Sec. 5.6. The performance of this scheme is evaluated in Appendix K and appears below.

Theorem 11 *While $R \lesssim J_{\mathrm{vox}} 2^{(M-1)\frac{J_{\mathrm{vox}}}{n}}$ where $J_{\mathrm{vox}} = \frac{n \cdot \min(\pi_i)}{\min\left(K_{\mathrm{d}}^{\mathrm{sp}}, 2K_{\mathrm{s}}^{\mathrm{sp}}+1\right)}$, the coding scheme from Sec. 5.5 applied to encode the approximation $\widehat{\widetilde{f_\pi^{\mathrm{s}}}}$ to $\widetilde{f_\pi^{\mathrm{s}}}$ using a discrete wavelet/surfprint dictionary achieves the following near-optimal asymptotic rate-distortion performance (within a logarithmic factor of the optimal performance of Theorem 2):*

$$\left\| \widetilde{f_\pi^{\mathrm{s}}} - \widehat{\widetilde{f_\pi^{\mathrm{s}}}} \right\|_{L_2}^2 \lesssim \left(\frac{\log R}{R} \right)^{\min\left(\frac{K_{\mathrm{d}}}{M-1}, \frac{2K_{\mathrm{s}}}{M}\right)}.$$

Again, a corollary follows naturally (see Appendix L for the proof).

Corollary 12 *While $R \lesssim J_{\mathrm{vox}} 2^{(M-1)\frac{J_{\mathrm{vox}}}{n}}$, the discrete approximation $\widehat{\widetilde{f_\pi^{\mathrm{s}}}}$ provides an approximation to f^{s} with the following rate-distortion performance:*

$$\left\| f^{\mathrm{s}} - \widehat{\widetilde{f_\pi^{\mathrm{s}}}} \right\|_{L_2}^2 \lesssim \left(\frac{\log R}{R} \right)^{\min\left(\frac{K_{\mathrm{d}}}{M-1}, \frac{2K_{\mathrm{s}}}{M}\right)}.$$

6.4 Discretization effects and varying sampling rates

We have proposed surflet algorithms for discrete data at sufficiently coarse scales. Unfortunately, this leaves open the question of how to represent such data at finer scales. In this section, we discuss one perspective on fine scale approximation that leads to a natural surflet coding strategy.

Consider again the class $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$. Section 6.2 provided an effective strategy for encoding elements of $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$ at sufficiently low rates (using surflets at scales $j \leq J_{\mathrm{vox}} = \frac{\min \pi_i}{K_{\mathrm{d}}}$).

Beyond scale J_{vox}, however, the voxelization effects dominate the resolution afforded by surflet approximations. To restore a balance, we suggest a coding strategy for finer scales based on the observation that $\mathcal{F}_{\text{C}}(M, K_{\text{d}}) \subset \mathcal{F}_{\text{C}}(M, K)$ for $K < K_{\text{d}}$. Surflet approximations on the class $\mathcal{F}_{\text{C}}(M, K)$ (tied to the smoothness K) have lower accuracy in general. As a result, $\widetilde{\mathcal{F}_{\text{C}}}(M, K)$ has a higher "breakdown rate" than $\widetilde{\mathcal{F}_{\text{C}}}(M, K_{\text{d}})$, and discrete surflets tailored for smoothness K will achieve the coding rate $O(R^{-K/(M-1)})$ up to scale $\frac{\min \pi_i}{K}$. While this may not be a worthwhile strategy before scale J_{vox}, it could be useful beyond scale J_{vox} and up to scale $\frac{\min \pi_i}{K}$. In fact, beyond that scale, we can *again* reduce K, obtaining a new breakdown rate and a finer scale to code (using lower-order surflets). This gives us a concrete strategy for coding $\widetilde{\mathcal{F}_{\text{C}}}(M, K_{\text{d}})$ at all scales, although our optimality arguments apply only up to scale J_{vox}. At scale j, we use surflets designed for smoothness $K_j = \min\left(K_{\text{d}}, \frac{\min(\pi_i)}{j}\right),\ 0 \le j \le \min(\pi_i)$. A surflet dictionary constructed using such scale-adaptive smoothness orders consists of relatively few elements at coarse scales (due to the low value of j in the quantization stepsize) and relatively few at fine scales (due to the decrease of K_j), but many elements at medium scales. This agrees with the following intuitive notions:

- The large block sizes at coarse scales do not provide sufficient resolution to warrant large dictionaries for approximation at these scales.
- The relatively small number of voxels in each block at very fine scales also means that a coder does not require large dictionaries in order to approximate blocks at such scales well.
- At medium scales where the block sizes are small enough to provide good resolution but large enough to contain many voxels, the dictionary contains many elements in order to provide good approximations.

Similar strategies can be proposed, of course, for the class $\widetilde{\mathcal{F}_{\text{S}}}(M, K_{\text{d}}, K_{\text{s}})$.

Finally we note that the interplay between the sampling rate (number of voxels) along the different dimensions and the critical approximation scale J_{vox} can impact the construction of multiscale source coders. As an example of the potential effect of this phenomenon in real-world applications, the sampling rate along the temporal dimension could be the determining factor when designing a surfprint-based video coder because this rate tends to be lower than the sampling rate along the spatial dimensions.

6.5 Simulation results

To demonstrate the potential for coding gains based on surflet representations, we perform the following numerical experiments in 2 and 3 dimensions.

6.5.1 2D coding

We start by coding elements of $\widetilde{\mathcal{F}_{\text{C}}}(M, K_{\text{d}})$ with $M = 2$ and $K_{\text{d}} = 3$. We generate 1024×1024 discretized versions of these images (that is, $\pi_1 = \pi_2 = 10$). Our two example images are shown in Figs. 6(a) and 7(a).

On each image we test three types of surflet dictionaries for encoding.

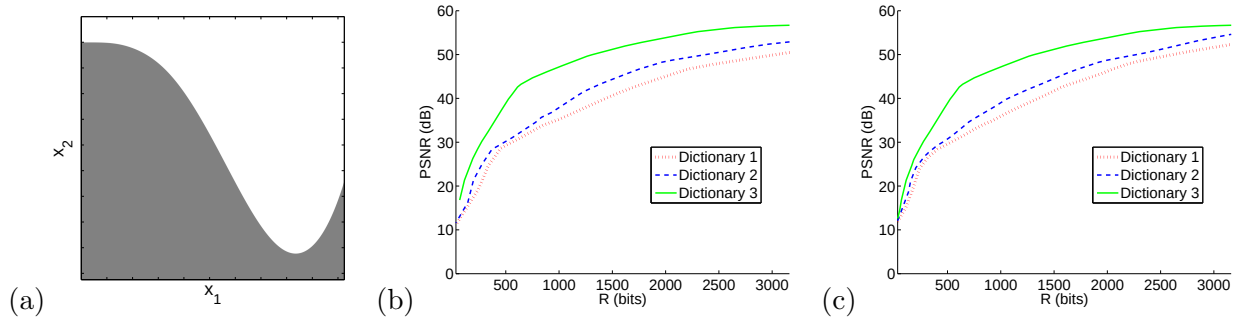


Figure 6: *(a) Test function $\widetilde{f^{\mathrm{c}}_{\pi}}$. (b) Rate-distortion performance for each dictionary (with the best fixed set of dictionary parameters). (c) Rate-distortion performance for each dictionary (selected using best convex hull in R/D plane over all dictionary parameters).*

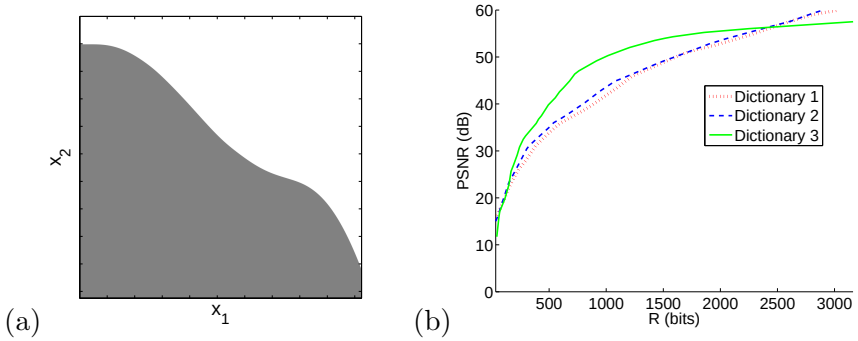


Figure 7: *(a) Test function $\widetilde{f^{\mathrm{c}}_{\pi}}$. (b) Rate-distortion performance for each dictionary (selected using best convex hull in R/D plane over all dictionary parameters).*

- Dictionary 1 uses wedgelets as implemented in our previous work [23, 29]. In this dictionary we do not use the quantization stepsizes as specified in (3). Rather, we use a quantization stepsize $\Delta_{\ell,j} \sim 2^{-(1-\ell)j}$. As a result, the quantized wedgelet dictionary has the same cardinality at each scale and is self-similar (simply a dyadic scaling of the dictionary at other scales).

- Dictionary 2 adapts with scale. Following the arguments of Sec. 6.4, at a given scale j, we use surflets tailored for smoothness $K_j = \min(2, \frac{\min \pi_i}{j}) = \min(2, \frac{10}{j})$. We use surflets of the appropriate polynomial order and quantize the polynomial coefficients analogous to (3); that is, $\Delta_{\ell,j} \sim 2^{-(K_j-\ell)j}$. The limitation $K_j \leq 2$ restricts our surflets to linear polynomials (wedgelets) for comparison with the first dictionary above.

- Dictionary 3 is a surflet dictionary that also adapts with scale. This dictionary is constructed similarly to the second, except that it is tailored to the actual smoothness of f^{c}: we set $K_j = \min(K_{\mathrm{d}}, \frac{\min \pi_i}{j}) = \min(K_{\mathrm{d}}, \frac{10}{j})$. This modification allows quadratic surflets to be used at coarse scales $0 \leq j \leq 5$, beyond which K_j again dictates that wedgelets are used.

For each dictionary, we must also specify the range of allowable polynomial coefficients and a constant multiplicative factor on each quantization stepsize. We optimize these parameters through simulation.

Table 1: *Surflet dictionary size at each scale (using the surflet parameters chosen to generate Fig. 6(b)). Our surflet dictionaries (2 and 3) adapt to scale, avoiding unnecessary precision at coarse and fine scales.*

Scale j	0	1	2	3	4	5	6	7	8	9
Dictionary 1	181631	181631	181631	181631	181631	181631	181631	181631	181631	181631
Dictionary 2	219	4143	62655	987903	987903	248191	62655	15967	4143	1119
Dictionary 3	357	14335	407719	12216207	6264455	248191	62655	15967	4143	1119

Our coding strategy for each dictionary uses a top-down prediction. Based on the prediction from a (previously coded) parent surflet, we partition the set of possible children surflets into two classes for entropy coding. A probability mass of ρ is distributed among the W surflets nearest the predicted surflet (measured using ℓ_2 distance), and a probability mass of $(1-\rho)$ is distributed among the rest to allow for robust encoding. We optimize the choice of W and ρ experimentally.

To find the discrete ℓ_2-best fit surflet to a given block, we use a coarse-to-fine iterative algorithm to search for the closest point along the manifold of possible surflets; we refer the reader to [30] for a detailed description of this algorithm. Based on the costs incurred by this coding scheme, we optimize the surflet tree pruning using a Lagrangian tradeoff parameter λ. We repeat the experiment for various values of λ.

Figure 6(b) shows what we judge to be the best R/D curve for each dictionary (Dictionary 1: dotted curve, 2: dashed curve, and 3: solid curve). Each curve is generated by sweeping λ but fixing one combination of polynomial parameters/constants. Over all simulations (all polynomial parameters/constants), we also take the convex hull over all points in the R/D plane. The results are plotted in Figs. 6(c) and 7(b).

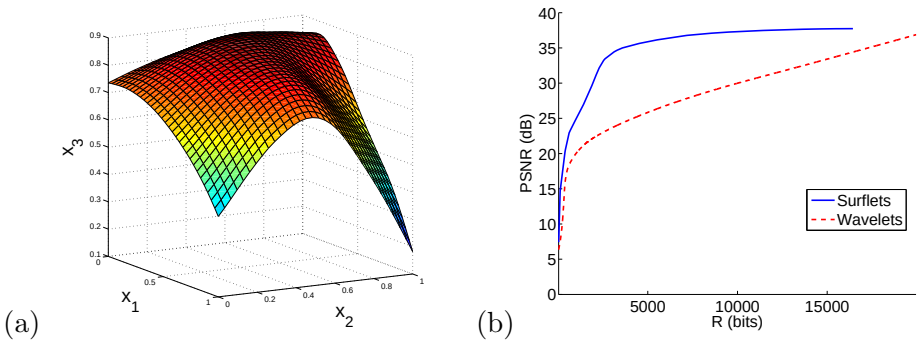
We see from the figures that Dictionary 2 outperforms Dictionary 1, requiring 0-20% fewer bits for an equivalent distortion (or improving PSNR by up to 4dB at a given bitrate). Both dictionaries use wedgelets — we conclude that the coding gain comes from the adaptivity through scale. Table 1 lists the number of admissible quantized surflets as a function of scale j for each of our three dictionaries.

We also see from the figures that Dictionary 3 often outperforms Dictionary 2, requiring 0-50% fewer bits for an equivalent distortion (or improving PSNR by up to 10dB at a given bitrate). Both dictionaries adapt to scale — we conclude that the coding gain comes from the quadratic surflets used at coarse scales (which are designed to exploit the actual smoothness $K_{\mathrm{d}} = 3$). Figure 4 compares two pruned surflet decompositions using Dictionaries 2 and 3. In this case, the quadratic dictionary offers comparable distortion using 40% fewer bits than the wedgelet dictionary.

6.5.2 3D coding

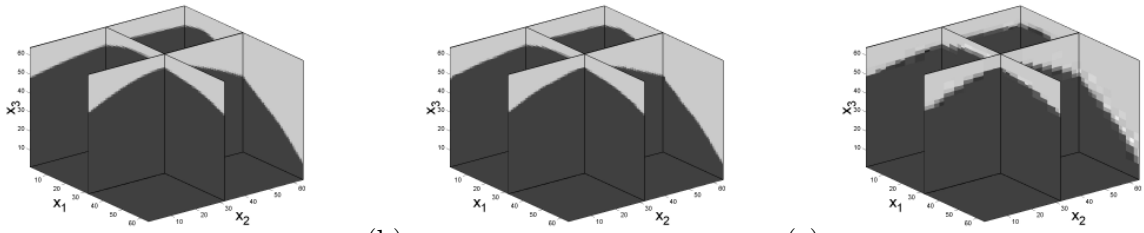
We now describe numerical experiments for coding elements of $\widetilde{\mathcal{F}_{\mathrm{C}}}(M, K_{\mathrm{d}})$ and $M = 3$. We generate $64 \times 64 \times 64$ discretized versions of these signals (that is, $\pi_i = 6$). Our two example discontinuities b^{c} are shown in Fig. 8(a) (for which $K_{\mathrm{d}} = 2$) and Fig. 10(a) (for which $K_{\mathrm{d}} = \infty$).

For these simulations we compare surflet coding (analogous to Dictionary 2 above, with $K_j =$



(a) (b)

Figure 8: *(a) Horizon b^c used to generate 3D test function $\widetilde{f^c_\pi}$. (b) Rate-distortion performance for surflet coding compared with wavelet coding.*



(a) (b) (c)

Figure 9: *Volumetric slices of 3D coded functions. (a) Original test function $\widetilde{f^c_\pi}$ from Fig. 8. (b) Surflet-coded function using 2540 bits; PSNR 33.22dB. (c) Wavelet-coded function using approximately 2540 bits; PSNR 23.08dB.*

$\min(2, \frac{6}{j})$) with wavelet coding. Our wavelet coding is based on a 3D Haar wavelet transform, which we threshold at a particular level (keeping the largest wavelet coefficients). For the purpose of the plots we assume (optimistically) that each significant wavelet coefficient was coded with zero distortion using only three bits per coefficient. We see from the figures that surflet coding significantly outperforms the wavelet approach, requiring up to 80% fewer bits than our aggressive wavelet estimate (or improving PSNR by up to 10dB a given bitrate). Figure 9 shows one set of coded results for the function in Fig. 8; at an equivalent bitrate, we see that surflets offer a significant improvement in PSNR and a dramatic reduction in ringing/blocking artifacts compared with wavelets. We also notice from Figs. 8 and 10, however, that at high bitrates the gains diminish relative to wavelets. We believe this is due to small errors made in the surflet estimates at fine scales using our current implementation of the manifold-based technique. Future work will focus on improved surflet estimation algorithms; however using even these suboptimal estimates we *still* see superior performance across a wide range of bitrates.

7 Conclusion

In this paper, we have studied the representation and compression of piecewise constant and piecewise smooth functions with smooth discontinuities. For both classes of functions, we determined

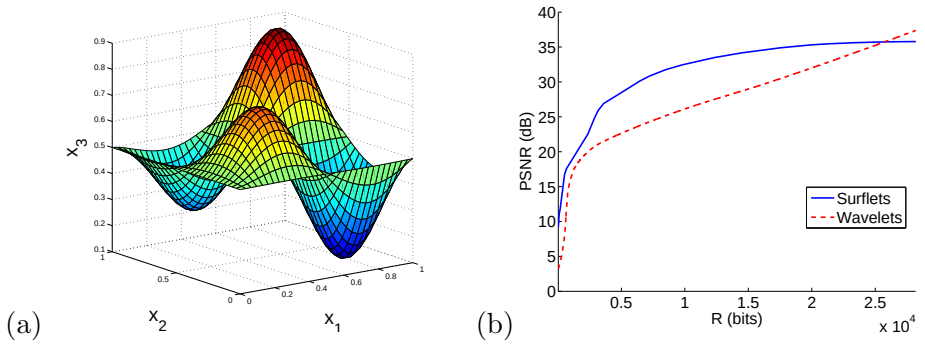


Figure 10: *(a) Horizon b^{c} used to generate 3D test function $\widetilde{f_{\pi}^{\mathrm{c}}}$. (b) Rate-distortion performance for surflet coding compared with wavelet coding.*

the metric entropy and then provided compression strategies based on multiresolution predictive coding in order to realize significant gains in rate-distortion performance. For piecewise constant functions, our surflet-based compression framework approximates and encodes such functions by assembling piecewise approximations over dyadic hypercubes at different scales, where each surflet approximant contains a (high-order) polynomial discontinuity that separates two constant regions. This surflet-based approach achieves the optimal approximation performance and the metric entropy bound. For piecewise smooth functions, we derived surfprints by combining surflets with wavelets. Our surfprint-based compression framework provides optimal approximation performance and near-optimal rate-distortion performance.

In addition, we extended our results for the continuous signal classes $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ and $\mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$ to their corresponding discrete function spaces. We provided asymptotic performance results for both discrete function spaces and related this asymptotic performance to the sampling rate and smoothness orders of the underlying functions and discontinuities. Our simulation results for 2D discrete piecewise constant functions demonstrate the coding gains achieved by using higher-order polynomials in the construction of surflet-based approximations and by defining surflet dictionaries based on scale-adaptive smoothness orders. Our 3D simulation results show that surflet-based approximations are vastly superior to wavelet-based methods over a large range of bitrates.

The insights that we gained, namely, in quantizing higher-order terms with lesser precision and using predictive coding to decrease bitrate, can be used to solve more sophisticated signal representation problems. In addition, our methods require knowledge only of the higher-dimensional function and not the smooth discontinuity. We conclude that surfprint-based representations will enable significant rate-distortion gains in realistic scenarios, as supported by our simulation results.

Future research will focus on the application of the approximation schemes presented in this paper to statistical estimation of higher dimensional signals containing arbitrary smooth discontinuities given noisy data (extending the piecewise constant $K_{\mathrm{d}} = 2$, $M = 2$ case treated in [8]). We would like to develop new representation schemes that provide optimal approximation performance for functions containing multiple intersecting discontinuities with high smoothness orders (for $M > 2$). We are also interested in studying practical applications of our coding schemes to the compression of natural images (addressing issues similar to those discussed in [23]), video, and light-field data. Such a study will involve tuning the various parameters related to the quantization

schemes, and the construction of the surflet and surfprint dictionaries. Additional work will also improve practical methods for fitting ℓ_2-best surflets, extending the methods described in [30].

Acknowledgements

Thanks to Justin Romberg, Ron DeVore, Minh Do, Mark Embree, and Rebecca Willett for enlightening discussions; Peter Jones for inspiring conversations on β-numbers; and Emmanuel Candès, David Donoho, and Albert Cohen for helpful questions.

A Proof of Theorem 1

Main idea: Let $\underline{f_N^{\mathrm{c}}}$ be an N-term approximant from any representation scheme that provides an approximation to a function $f^{\mathrm{c}} \in \mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ using N terms. We show that we can construct a Horizon-class function from $\underline{f_N^{\mathrm{c}}}$ with the same asymptotic rate as this approximation scheme. As a result, it follows that we only need to consider Horizon-class approximation schemes in establishing a bound on the optimal asymptotic approximation performance for $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$. (There cannot exist a scheme using non-Horizon approximants that performs better asymptotically.) This connection allows us to directly apply the approximation and metric entropy results pertaining to the $(M-1)$-dimensional discontinuity.

Approximation: Let $\mathbf{x} \in [0,1]^M$ and let $\mathbf{y}$ denote the first $M-1$ elements of $\mathbf{x}$ (as defined in Sec. 2.2). Define a function $\underline{f}$ such that

$$\underline{f}(\mathbf{x}) = \begin{cases} 1, & \underline{f_N^{\mathrm{c}}}(\mathbf{x}) > 0.5 \\ 0, & \text{otherwise.} \end{cases}$$

Considering the four cases of f^{c} being 0 or 1 and $\underline{f}$ being 0 or 1, we have

$$\left\| f^{\mathrm{c}} - \underline{f} \right\|_{L_2}^2 \leq 4 \cdot \left\| f^{\mathrm{c}} - \underline{f_N^{\mathrm{c}}} \right\|_{L_2}^2 . \tag{11}$$

Now we construct a Horizon-class function from $\underline{f}$. Let $\widehat{b_N^{\mathrm{c}}}$ be an $(M-1)$-dimensional function defined as

$$\widehat{b_N^{\mathrm{c}}}(\mathbf{y}) = 1 - \int_0^1 \underline{f}(\mathbf{y}, x_M) dx_M.$$

Finally, let $\widehat{f_N^{\mathrm{c}}}$ be a Horizon-class function defined by the $(M-1)$-dimensional discontinuity $\widehat{b_N^{\mathrm{c}}}$:

$$\widehat{f_N^{\mathrm{c}}}(\mathbf{x}) = \begin{cases} 1, & x_M \geq \widehat{b_N^{\mathrm{c}}}(\mathbf{y}) \\ 0, & x_M < \widehat{b_N^{\mathrm{c}}}(\mathbf{y}). \end{cases}$$

Again, considering the four cases of f^{c} being 0 or 1 and $\widehat{f_N^{\mathrm{c}}}$ being 0 or 1, we have

$$\left\| f^{\mathrm{c}} - \widehat{f_N^{\mathrm{c}}} \right\|_{L_2}^2 \leq \left\| f^{\mathrm{c}} - \underline{f} \right\|_{L_2}^2 ,$$

and using (11) we conclude that

$$\left\| f^{\mathrm{c}} - \widehat{f_N^{\mathrm{c}}} \right\|_{L_2}^2 \leq 4 \cdot \left\| f^{\mathrm{c}} - \underline{f_N^{\mathrm{c}}} \right\|_{L_2}^2 . \tag{12}$$

This result shows that the approximation performance of any scheme that approximates f^{c} is bounded below by the approximation performance of a corresponding Horizon-class representation scheme.

Because $\widehat{f_N^{\mathrm{c}}}$ is a Horizon-class function,

$$\left\| f^{\mathrm{c}} - \widehat{f_N^{\mathrm{c}}} \right\|_{L_2}^2 \leq \left\| b^{\mathrm{c}} - \widehat{b_N^{\mathrm{c}}} \right\|_{L_1} , \tag{13}$$

where b^{c} is the $\mathcal{C}^{K_{\mathrm{d}}}$ discontinuity in f^{c} and $\widehat{b_N^{\mathrm{c}}}$, the implicit estimate to b^{c}, is the Horizon discontinuity in $\widehat{f_N^{\mathrm{c}}}$. From the work of Cohen et al. [14] regarding optimal approximation rates, the optimal approximation rate for the $(M-1)$-dimensional $\mathcal{C}^{K_{\mathrm{d}}}$ class of functions is

$$\left\| b^{\mathrm{c}} - \widehat{b_N^{\mathrm{c}}} \right\|_{L_1} \lesssim \left(\frac{1}{N} \right)^{\frac{K_{\mathrm{d}}}{M-1}} . \tag{14}$$

Combining (12), (13), and (14), we have an upper-bound on achievable approximation performance for $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$:

$$\left\| f^{\mathrm{c}} - \widehat{f_N^{\mathrm{c}}} \right\|_{L_2}^2 \lesssim \left(\frac{1}{N} \right)^{\frac{K_{\mathrm{d}}}{M-1}} . \tag{15}$$

However, (13) is satisfied with equality when both b^{c} and $\widehat{b_N^{\mathrm{c}}}$ are completely contained inside the unit hypercube (i.e., $0 \leq b^{\mathrm{c}}(\mathbf{y}), \widehat{b_N^{\mathrm{c}}}(\mathbf{y}) \leq 1$), and we also know that (14) provides the *optimal* approximation rate for the $(M-1)$-dimensional $\mathcal{C}^{K_{\mathrm{d}}}$ class of functions. Thus, (15) provides the optimal approximation performance that could be achieved for *every* $f^{\mathrm{c}} \in \mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$.

Rate distortion: To find the optimal rate-distortion performance for $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$, a similar argument could be made using the work of Clements [20] (extending Kolmogorov and Tihomirov [19]) regarding metric entropy. It follows from these papers that the optimal asymptotic rate-distortion performance for the $(M-1)$-dimensional $\mathcal{C}^{K_{\mathrm{d}}}$ class of functions is

$$\left\| b^{\mathrm{c}} - \widehat{b_R^{\mathrm{c}}} \right\|_{L_1} \lesssim \left(\frac{1}{R} \right)^{\frac{K_{\mathrm{d}}}{M-1}} .$$

□

B Proof of Theorem 2

Let f^{s} be defined by the $(M-1)$-dimensional $\mathcal{C}^{K_{\mathrm{d}}}$ discontinuity b^{s} separating two M-dimensional $\mathcal{C}^{K_{\mathrm{s}}}$ functions $g_1^{\mathrm{s}}, g_2^{\mathrm{s}}$. We first establish a lower bound on the optimal approximation rate with respect to the squared-L_2 distortion measure for $\mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$. We note that both the space of M-dimensional uniformly $\mathcal{C}^{K_{\mathrm{s}}}$-smooth functions and $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$ are subsets of $\mathcal{F}_{\mathrm{S}}(M, K_{\mathrm{d}}, K_{\mathrm{s}})$. Cohen et al. [14] show that the optimal approximation decay rate for the space of M-dimensional

uniformly $\mathcal{C}^{K_\mathrm{s}}$-smooth functions is $\left(\frac{1}{N}\right)^{\frac{2K_\mathrm{s}}{M}}$, while Theorem 1 proves that the optimal approximation rate for $\mathcal{F}_\mathrm{C}(M, K_\mathrm{d})$ is $\left(\frac{1}{N}\right)^{\frac{K_\mathrm{d}}{M-1}}$. Therefore, the optimal approximation rate for $\mathcal{F}_\mathrm{S}(M, K_\mathrm{d}, K_\mathrm{s})$ is bounded below by $\left(\frac{1}{N}\right)^{\min\left(\frac{K_\mathrm{d}}{M-1}, \frac{2K_\mathrm{s}}{M}\right)}$.

We now prove that this lower bound can be achieved, thus establishing the optimal approximation rate for $\mathcal{F}_\mathrm{S}(M, K_\mathrm{d}, K_\mathrm{s})$. We assume that explicit information about b^s and $g_1^\mathrm{s}, g_2^\mathrm{s}$ is provided by an external "oracle". Given such information about the M-dimensional $\mathcal{C}^{K_\mathrm{s}}$ functions, one could use a wavelet-based approximation scheme [14] to achieve the optimal approximation rate $\left(\frac{1}{N}\right)^{\frac{2K_\mathrm{s}}{M}}$ for such functions. Next, one could use a similar wavelet-based approach in $M-1$ dimensions to represent b^s with the optimal approximation rate of $\left(\frac{1}{N}\right)^{\frac{K_\mathrm{d}}{M-1}}$. Thus, we have provided the optimal approximation rate for every function $f^\mathrm{s} \in \mathcal{F}_\mathrm{S}(M, K_\mathrm{d}, K_\mathrm{s})$. (The assumption about availability of explicit information about b^s enables to prove the existence of efficient approximations for f^s.) Finally, given the results of Clements [20] (extending Kolmogorov and Tihomirov [19]), the optimal rate-distortion performance for $\mathcal{F}_\mathrm{S}(M, K_\mathrm{d}, K_\mathrm{s})$ can be derived similarly. □

C Proof of Theorem 3

Consider a candidate surflet decomposition grown fully up to level J but pruned back in regions away from the discontinuity to consolidate nodes that are entirely 0- or 1-valued. This surflet decomposition then consists of the following leaf nodes:

- dyadic hypercubes at level J through which the singularity b passes, and which are decorated with a surflet; and
- dyadic hypercubes at various levels through which the singularity b does not pass, and which are all-zero or all-one.

We establish the asymptotic approximation rate for this candidate decomposition. Because this configuration is among the options available to the approximation rate optimized tree-pruning in Sec. 4.3, this provides an upper bound on the asymptotic approximation rate of the algorithm.

Distortion: First we establish a bound on the distortion in such a decomposition. We assume quantized Taylor surflets for this analysis; this provides an upper bound for the distortion of native L_2-best surflets as well (since native L_2-best surflets are chosen from a dictionary that includes the quantized Taylor surflets). In fact, the behavior of the upper bound will also hold for extended L_2-best surflets, but with slightly larger constants. Let X_J be a dyadic hypercube at level J, and let $\mathbf{y}_\mathrm{ep}$ be its expansion point. Using Taylor's theorem (with $D = M-1, d = b^\mathrm{c}, K = K_\mathrm{d}, r = r_\mathrm{d}, \alpha = \alpha_\mathrm{d}$ in (1)), we construct a polynomial approximation of the discontinuity b^c using the Taylor surflet

$s^T(X_J; p; \cdot)$ as follows. For each $\mathbf{y} \in Y_J$

$$\begin{aligned} p(\mathbf{y}) = p(\mathbf{y}_{\text{ep}} + \mathbf{h}) &= [b^{\text{c}}(\mathbf{y}_{\text{ep}}) + c_0 \cdot 2^{-K_{\text{d}} J}] \\ &+ \frac{1}{1!} \sum_{i_1=1}^{M-1} [b^{\text{c}}_{y_{i_1}}(\mathbf{y}_{\text{ep}}) + c_{1,i_1} \cdot 2^{-(K_{\text{d}}-1)J}] \cdot h_{i_1} \\ &+ \frac{1}{2!} \sum_{i_1,i_2=1}^{M-1} [b^{\text{c}}_{y_{i_1},y_{i_2}}(\mathbf{y}_{\text{ep}}) + c_{2,i_1,i_2} \cdot 2^{-(K_{\text{d}}-2)J}] \cdot h_{i_1} h_{i_2} + \cdots \\ &+ \frac{1}{r_{\text{d}}!} \sum_{i_1,\cdots,i_{r_{\text{d}}}=1}^{M-1} [b^{\text{c}}_{y_{i_1},\cdots,y_{i_{r_{\text{d}}}}}(\mathbf{y}_{\text{ep}}) + c_{r_{\text{d}},i_1,\cdots,i_{r_{\text{d}}}} \cdot 2^{-\alpha_{\text{d}} J}] \cdot h_{i_1} \cdots h_{i_{r_{\text{d}}}}, \quad (16) \end{aligned}$$

where each constant $c_{\ell,i_1,\cdots,i_\ell}$ depends on $b^{\text{c}}_{y_{i_1},\cdots,y_{i_\ell}}(\mathbf{y}_{\text{ep}})$, and $|c_{\ell,i_1,\cdots,i_\ell}| \leq \frac{1}{2}$. The surflet polynomial discontinuity p is constructed by using quantized values of the derivatives of b^{c} evaluated at $\mathbf{y}_{\text{ep}}$ as the polynomial coefficients. The set of all such $s^T(X_J; p; \cdot)$'s is precisely $\mathcal{M}(J)$. From Taylor's theorem (1) and (16), we have that the X_J-clipped L_1 distance between b^{c} and p is

$$\overline{L_1}(b^{\text{c}}, p) \leq C_3 \cdot 2^{-K_{\text{d}} J - (M-1)J}. \quad (17)$$

Thus we have that the squared-L_2 error between f^{c} and $s^T(X_J; p; \cdot)$ over X_J is

$$\left\| f^{\text{c}} - s^T(X_J; p; \cdot) \right\|^2_{L_2(X_J)} \leq C_3 \cdot 2^{-K_{\text{d}} J - (M-1)J}. \quad (18)$$

We construct an approximation $\widehat{f^{\text{c}}_J}$ to f^{c} at scale J by tiling together all the surflets $s^T(X_J; p; \cdot)$ (where the surflet polynomial p differs from one hypercube to another). Let $N_{j,\text{SL}}$ denote the number of nodes at level j through which the discontinuity b passes. Due to the bounded curvature of b^{c}, $N_{j,\text{SL}} \lesssim 2^{(M-1)j}$. Therefore, we have that the total distortion is

$$\left\| f^{\text{c}} - \widehat{f^{\text{c}}_J} \right\|^2_{L_2} \leq C_4 \cdot 2^{-K_{\text{d}} J}. \quad (19)$$

Number of terms: Next we establish a bound on the number of surflets required to encode this decomposition (using either quantized Taylor surflets or L_2-best surflets). We know from above that due to the bounded curvature of b^{c}, the number of nodes at level j through which the discontinuity b^{c} passes is given by $N_{j,\text{SL}} \lesssim 2^{(M-1)j}$. Let $N_{j,\text{ZO}}$ be the number of all-zero and all-one nodes in the pruned decomposition at level j:

$$N_{j,\text{ZO}} \leq 2^M \cdot N_{j-1,\text{SL}} \leq 2^M \cdot C_5 \cdot 2^{(M-1)(j-1)} \leq C_6 \cdot 2^{(M-1)j}.$$

Thus, the number of terms required in the approximation is given by

$$N \leq \sum_{j=0}^{J} N_{j,\text{SL}} + N_{j,\text{ZO}} \leq C_7 \cdot 2^{(M-1)J}. \quad (20)$$

Finally, we combine (19) and (20) to obtain the result. □

D Proof of Theorem 4

As in Theorem 3, we consider a candidate surflet decomposition grown fully up to level J but pruned back in regions away from the discontinuity. This surflet decomposition then consists of the following nodes:

- leaf nodes at level J, which are decorated with a surflet;
- leaf nodes at various levels, which are all-zero or all-one; and
- internal nodes at various levels, which are decorated with surflets.

Since the leaf nodes are used to construct the approximation, we may use the distortion bound (19) from Theorem 3. This bound holds for both quantized Taylor surflets and L_2-best surflets.

Number of bits: We establish a bound on the bitrate required to encode this decomposition (using either quantized Taylor surflets or L_2-best surflets). There are three contributions to the bitrate:

- To encode the structure (topology) of the pruned tree indicating the locations of the leaf nodes, we can use one bit for each node in the tree [31]. Using (20), we have

$$R_1 \leq \sum_{j=0}^{J} N_{j,\mathrm{SL}} + N_{j,\mathrm{ZO}} \leq C_7 \cdot 2^{(M-1)J}. \tag{21}$$

- For each leaf node that is all-zero or all-one we can use a constant number of bits to specify the homogeneous nature of the node and the constant value (0 or 1). We have

$$R_2 \leq \sum_{j=0}^{J} N_{j,\mathrm{ZO}} \leq C_8 \cdot 2^{(M-1)J}. \tag{22}$$

- For each leaf node at scale J labeled with a surflet, we must encode the quantized surflet parameters. For a surflet coefficient at scale J of order $\ell \in \{0, \cdots, r_\mathrm{d}\}$, the number of bits required per coefficient is $O((K_\mathrm{d} - \ell)J)$, and the number of such coefficients is $O((M-1)^\ell)$. Hence, the total number of bits required to encode each surflet is $O(J \sum_{\ell=0}^{r_\mathrm{d}} (K_\mathrm{d} - \ell)(M-1)^\ell) = O(J)$. (Note that our order term describes the scaling with J.) Therefore, we have that

$$R_3 \leq N_{J,\mathrm{SL}} \cdot O(J) \leq C_9 \cdot J \cdot 2^{(M-1)J}. \tag{23}$$

Combining (21), (22), and (23), the total bitrate $R(f^\mathrm{c}, \widehat{f_J^\mathrm{c}})$ required to describe the surflet decomposition satisfies

$$R(f^\mathrm{c}, \widehat{f_J^\mathrm{c}}) = R_1 + R_2 + R_3 \leq C_{10} \cdot J \cdot 2^{(M-1)J}.$$

We conclude the proof by combining this result with (19). □

E Proof of Theorem 5

The proof consists of three steps. First, we show that the *L-extended L_2-best surflet* fits to f^{c} over a hypercube X_j and one of its children X_{j+1} share a non-trivial common X_{j+1}^L-active region. Second, we use this fact to show that the surflet polynomial coefficients of these *L-extended L_2-best surflet* fits are similar in the case where the common X_{j+1}^L-active region is aligned with the center of the hypercube X_{j+1}. Finally, we extend the second step to show that the surflet polynomial coefficients of the *L-extended L_2-best surflet* fits are similar regardless of the exact location of the common X_{j+1}^L-active region. We combine these steps to prove that the surflet polynomial coefficients at scale $j+1$ can be encoded using a constant number of bits given the surflet polynomial coefficients at scale j.

Surflet fits over X_j and X_{j+1} share non-trivial active region: Assume that the discontinuity b^{c} passes through X_{j+1}. (Note that if b^{c} passes through X_j but not through X_{j+1}, our coder uses an all-0 or all-1 surflet to approximate f^{c} over X_{j+1}; checking for such an occurrence does not require explicit knowledge of b^{c}.) From (17), we have that the *L-extended L_2-best surflet* fits $s^L(X_j; p_j; \cdot)$ and $s^L(X_{j+1}; p_{j+1}; \cdot)$ obey the following relations:

$$\overline{L_1}(b^{\mathrm{c}}, p_j) \leq C_3 \cdot 2^{-(K_{\mathrm{d}}+M-1)j} \tag{24}$$

and

$$\overline{L_1}(b^{\mathrm{c}}, p_{j+1}) \leq C_3 \cdot 2^{-(K_{\mathrm{d}}+M-1)(j+1)}. \tag{25}$$

Since b^{c} passes through X_{j+1}, it is clear that b^{c} will have a nontrivial $(M-1)$-dimensional X_{j+1}^L-active region. One can also check that the surflet polynomials p_j and p_{j+1} have similar X_{j+1}^L-active regions. In particular, there exists an $(M-1)$-dimensional hypercube region A_{j+1} that is contained in the X_{j+1}^L-active regions of b^{c}, p_j, and p_{j+1}, with sidelength$(A_{j+1}) = C_{11} \cdot 2^{-(j+1)}$ in each dimension with the constant C_{11} independent of j; otherwise the bounds (24) and (25) could not hold with b^{c} passing through X_{j+1}. Hence $\mathrm{vol}(A_{j+1}) = (C_{11})^{M-1} \cdot 2^{-(M-1)(j+1)}$, where the constant C_{11} depends on the universal derivative bound for functions in $\mathcal{F}_{\mathrm{C}}(M, K_{\mathrm{d}})$, the extension L, and the dimension of the problem M. We emphasize here that the notion of L-extensions of hypercubes is the key reason that surflet fits at successive scales share such a non-trivial common X_{j+1}^L-active region.

We now restrict the domain of consideration of p_j to the L-extended domain corresponding to the hypercube X_{j+1} at scale $j+1$. We denote the resulting polynomial by p_j^{j+1}. Thus, using (24), (25), and the triangle inequality, the L_1 distance between p_j^{j+1} and p_{j+1} over the $(M-1)$-dimensional active region A_{j+1} is bounded by

$$\|p_j^{j+1} - p_{j+1}\|_{L_1(A_{j+1})} \leq C_{12} \cdot 2^{-(K_{\mathrm{d}}+M-1)(j+1)}, \tag{26}$$

with the constant C_{12} independent of j.

Surflet polynomial coefficients are similar when A_{j+1} is centered with respect to X_{j+1}: We first assume for simplicity that $A_{j+1} = [-\frac{C_{11}}{2} \cdot 2^{-(j+1)}, \frac{C_{11}}{2} \cdot 2^{-(j+1)}]^{M-1}$. In order to relate the similarity (26) between p_j and p_{j+1} in the L_1 sense to their polynomial coefficients, we present the following lemma.

Lemma 1 [32, pp. 72-73] *Let $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_D\}$ be a set of linearly independent vectors. Then there exists $C > 0$ such that for any collection of coefficients $\gamma = \{\gamma_i\}_{i=1}^D$,*

$$|\gamma_i| \leq C \cdot \|\boldsymbol{v}\|, \ \forall i \in \{1, \cdots, D\},$$

where $\boldsymbol{v} = \sum_{i=1}^D \gamma_i \boldsymbol{v}_i$. (For any norm, such a constant exists.)

In order to employ the lemma, we define a monomial basis that contains all monomials of the form $y_1^{i_1} \cdot y_2^{i_2} \cdots y_{M-1}^{i_{M-1}}$, where $i_1 + i_2 + \cdots + i_{M-1} \in \{0, \ldots, r_{\mathrm{d}}\}$. We denote by $\mathbf{v}_{i,\ell}$ a monomial basis element of order $i_1 + i_2 + \cdots + i_{M-1} = \ell$, with i as an index (i specifies the powers in the basis monomial). At level $j+1$, the domain of each $\mathbf{v}_{i,\ell}$ is restricted to A_{j+1}. We express p_{j+1} and p_j^{j+1} as polynomials in the vector space spanned by the monomial basis:

$$p_j^{j+1} = \sum_{i,\ell} a_{i,\ell,p_j^{j+1}}^{j+1} \mathbf{v}_{i,\ell}$$

and

$$p_{j+1} = \sum_{i,\ell} a_{i,\ell,p_{j+1}}^{j+1} \mathbf{v}_{i,\ell}.$$

Now, we define an error vector

$$\mathbf{e}^{j+1} = p_j^{j+1} - p_{j+1}.$$

We define $a_{i,\ell}^{j+1} = a_{i,\ell,p_j^{j+1}}^{j+1} - a_{i,\ell,p_{j+1}}^{j+1}$, and so we have $\mathbf{e}^{j+1} = \sum_{i,\ell} a_{i,\ell}^{j+1} \mathbf{v}_{i,\ell}$. Using (26), over a subdomain A_{j+1} of volume $(C_{11})^{M-1} \cdot 2^{-(M-1)(j+1)}$:

$$\|\mathbf{e}^{j+1}\|_{L_1(A_{j+1})} \leq C_{12} \cdot 2^{-(K_{\mathrm{d}}+M-1)(j+1)}. \tag{27}$$

To complete the proof, we must show that the L_2-best surflet polynomial coefficients at scale $j+1$ can be encoded using a constant number of bits given the surflet polynomial coefficients at scale j. To do so, note that the quantization bin-size (3) of an ℓ'th order coefficient at scale $j+1$ is $2^{-(K_{\mathrm{d}}-\ell)(j+1)}$. Consequently, it suffices to show that

$$|a_{i,\ell}^{j+1}| \leq C_{13} \cdot 2^{-(K_{\mathrm{d}}-\ell)(j+1)} \tag{28}$$

with C_{13} independent of j. Such a bound on $|a_{i,\ell}^{j+1}|$ would imply that the L_2-best surflet coefficients at scale $j+1$ can be encoded using roughly $\log_2(2C_{13}+1)$ (*constant*) bits, given the L_2-best surflet coefficients encoded at scale j. We now proceed to establish (28) in the case where $A_{j+1} = [-\frac{C_{11}}{2} \cdot 2^{-(j+1)}, \frac{C_{11}}{2} \cdot 2^{-(j+1)}]^{M-1}$; in the third part of the proof we will adjust our arguments to establish (28) for arbitrary A_{j+1}.

Since we only have bounds on the L_1 distance between p_{j+1} and p_j^{j+1} over A_{j+1}, we restrict our attention to this shared active region. Normalizing the basis vectors $\mathbf{v}_{i,\ell}$ with respect to the domain A_{j+1}, we have that

$$\mathbf{e}^{A_{j+1}} = \sum_{i,\ell} c_{i,\ell}^{A_{j+1}} \mathbf{w}_{i,\ell}^{A_{j+1}},$$

where $\mathbf{w}_{i,\ell}^{A_{j+1}} = \mathbf{v}_{i,\ell}/\|\mathbf{v}_{i,\ell}\|_{L_1(A_{j+1})}$, $\|\mathbf{v}_{i,\ell}\|_{L_1(A_{j+1})} = \int_{A_{j+1}} |\mathbf{v}_{i,\ell}|$, and $c_{i,\ell}^{A_{j+1}} = a_{i,\ell}^{j+1} \cdot \|\mathbf{v}_{i,\ell}\|_{L_1(A_{j+1})}$. Let $\mathbf{v}_{i,\ell} = y_1^{i_1} \cdot y_2^{i_2} \cdots y_{M-1}^{i_{M-1}}$ be a basis monomial with $i_1 + i_2 + \cdots + i_{M-1} = \ell$. From the definition of the $\|\cdot\|_{L_1(A_{j+1})}$ norm, we have that

$$\begin{aligned} \|\mathbf{v}_{i,\ell}\|_{L_1(A_{j+1})} &= \int_{\mathbf{y} \in [-\frac{C_{11}}{2} \cdot 2^{-(j+1)}, \frac{C_{11}}{2} \cdot 2^{-(j+1)}]^{M-1}} |y_1^{i_1} \cdots y_{M-1}^{i_{M-1}}| \cdot dy_1 \cdots dy_{M-1} \\ &= 2^{-\ell} \cdot C_{11}^{\ell+M-1} \cdot \frac{2^{-\ell(j+1)} \cdot 2^{-(M-1)(j+1)}}{(i_1+1)\cdots(i_{M-1}+1)}. \end{aligned} \tag{29}$$

Note that C_{11} is independent of j. Because the basis vectors $\mathbf{w}_{i,\ell}^{A_{j+1}}$ are linearly independent over A_{j+1}, we know from the lemma that there exists a $C(j+1)$ such that

$$|c_{i,\ell}^{A_{j+1}}| \leq C(j+1) \cdot \|\mathbf{e}^{A_{j+1}}\|_{L_1(A_{j+1})}. \tag{30}$$

We need to show that $C(j+1)$ is actually independent of j. This would allow us to conclude (28) for $A_{j+1} = [-\frac{C_{11}}{2} \cdot 2^{-(j+1)}, \frac{C_{11}}{2} \cdot 2^{-(j+1)}]^{M-1}$ because

$$\begin{aligned} |a_{i,\ell}^{j+1}| &= \left| \frac{c_{i,\ell}^{A_{j+1}}}{\|\mathbf{v}_{i,\ell}\|_{L_1(A_{j+1})}} \right| \\ &= \left| \frac{c_{i,\ell}^{A_{j+1}} \cdot (i_1+1)\cdots(i_{M-1}+1)}{2^{-\ell} \cdot C_{11}^{\ell+M-1} \cdot 2^{-\ell(j+1)} \cdot 2^{-(M-1)(j+1)}} \right| \\ &\leq \frac{C(j+1) \cdot C_{12} \cdot (i_1+1)\cdots(i_{M-1}+1)}{2^{-\ell} \cdot C_{11}^{\ell+M-1}} \cdot 2^{-(K_{\mathrm{d}}-\ell)(j+1)}, \end{aligned} \tag{31}$$

where we obtain the second equality from (29), and the inequality from (27) and (30). Indeed, if $C(j+1)$ is independent of j we note from (31) that we could set $C_{13} = [C(j+1) \cdot C_{12} \cdot (i_1 + 1)\cdots(i_{M-1}+1)]/[2^{-\ell} \cdot C_{11}^{\ell+M-1}]$ in (28).

We let $\xi \in [-\frac{C_{11}}{2}, \frac{C_{11}}{2}]^{M-1}$ denote the "relative position" within the active hypercube region of a surflet. For *any* level j_1 with $A_{j_1} = [-\frac{C_{11}}{2} \cdot 2^{-j_1}, \frac{C_{11}}{2} \cdot 2^{-j_1}]^{M-1}$,

$$\frac{\mathbf{w}_{i,\ell}^{A_{j+1}}(\xi \cdot 2^{-(j+1)})}{\mathbf{w}_{i',\ell'}^{A_{j+1}}(\xi \cdot 2^{-(j+1)})} = \frac{\mathbf{w}_{i,\ell}^{A_{j_1}}(\xi \cdot 2^{-j_1})}{\mathbf{w}_{i',\ell'}^{A_{j_1}}(\xi \cdot 2^{-j_1})}$$

from (29). Setting $\ell' = 0$, we have that

$$\mathbf{w}_{i,\ell}^{A_{j+1}}(\xi \cdot 2^{-(j+1)}) = 2^{-(M-1)(j_1-j-1)} \cdot \mathbf{w}_{i,\ell}^{A_{j_1}}(\xi \cdot 2^{-j_1}). \tag{32}$$

Thus, we can construct a vector $\mathbf{e}^{A_{j_1}}$ at level j_1 using the *same* coefficients $c_{i,\ell}^{A_{j+1}}$, i.e., $c_{i,\ell}^{A_{j_1}} = c_{i,\ell}^{A_{j+1}}$

so that

$$\begin{aligned}
\|\mathbf{e}^{A_{j+1}}\|_{L_1(A_{j+1})} &= \int_{\mathbf{y}\in A_{j+1}} \left| \sum_{i,\ell} c_{i,\ell}^{A_{j+1}} \mathbf{w}_{i,\ell}^{A_{j+1}}(\mathbf{y}) \right| dy_1 \cdots dy_{M-1} \\
&\quad (\xi_i = y_i \cdot 2^{(j+1)}) \\
&= \int_{\xi\in[-\frac{C_{11}}{2},\frac{C_{11}}{2}]^{M-1}} \left| \sum_{i,\ell} c_{i,\ell}^{A_{j+1}} \mathbf{w}_{i,\ell}^{A_{j+1}}(\xi \cdot 2^{-(j+1)}) \right| \cdot 2^{-(M-1)(j+1)} d\xi_1 \cdots d\xi_{M-1} \\
&= \int_{\xi\in[-\frac{C_{11}}{2},\frac{C_{11}}{2}]^{M-1}} \left| \sum_{i,\ell} c_{i,\ell}^{A_{j+1}} \mathbf{w}_{i,\ell}^{A_{j_1}}(\xi \cdot 2^{-j_1}) \right| \cdot 2^{-(M-1)j_1} d\xi_1 \cdots d\xi_{M-1} \\
&\quad (y_i = \xi_i \cdot 2^{-j_1};\ c_{i,\ell}^{A_{j_1}} = c_{i,\ell}^{A_{j+1}}) \\
&= \int_{\mathbf{y}\in A_{j_1}} \left| \sum_{i,\ell} c_{i,\ell}^{A_{j_1}} \mathbf{w}_{i,\ell}^{A_{j_1}}(\mathbf{y}) \right| dy_1 \cdots dy_{M-1} \\
&= \|\mathbf{e}^{A_{j_1}}\|_{L_1(A_{j_1})},
\end{aligned}$$

where we get the third equality from (32). Since the coefficients at the two levels, $j+1$ and j_1, are the same, we can set $C_{14} = C(j+1)$ in (30)

$$|c_{i,\ell}^{A_{j_1}}| \leq C_{14} \cdot \|\mathbf{e}^{A_{j_1}}\|_{L_1(A_{j_1})}.$$

In this manner, one can show that

$$|c_{i,\ell}^{A_{j+1}}| \leq C_{14} \cdot \|\mathbf{e}^{A_{j+1}}\|_{L_1(A_{j+1})} \tag{33}$$

is true for for all j (because j_1 is arbitrary), and hence C_{14} is independent of j. Switching back to the original coefficients $a_{i,\ell}^{j+1}$, we have that

$$|a_{i,\ell}^{j+1}| \leq C_{13} \cdot 2^{-(K_{\mathrm{d}}-\ell)(j+1)}, \tag{34}$$

following the logic in (31), with C_{13} independent of j.

Surflet polynomial coefficients are similar independent of the specific location of A_{j+1}: We now suppose $A_{j+1} = [-\frac{C_{11}}{2} \cdot 2^{-(j+1)} + \eta_1, \frac{C_{11}}{2} \cdot 2^{-(j+1)} + \eta_1] \times \cdots \times [-\frac{C_{11}}{2} \cdot 2^{-(j+1)} + \eta_{M-1}, \frac{C_{11}}{2} \cdot 2^{-(j+1)} + \eta_{M-1}]$, where η denotes the "center" of A_{j+1}. As in the previous step, let $\mathbf{e}^{j+1} = \sum_{i,\ell} a_{i,\ell}^{j+1} \mathbf{v}_{i,\ell}$. Suppose that we transform the basis vectors $\mathbf{v}_{i,\ell}$ to be centered around η rather than around $\mathbf{0}$. The transformed coefficients would then satisfy the bound in (34). To make this point precise, let $\mathbf{v}_{i,\ell}^{A_{j+1}}(\mathbf{y}) = \mathbf{v}_{i,\ell}(\mathbf{y} - \eta) = (y_1 - \eta_1)^{i_1} \cdots (y_{M-1} - \eta_{M-1})^{i_{M-1}}$, and let $a_{i,\ell}^{A_{j+1}}$ denote the transformed coefficients of the error vector so that $\mathbf{e}^{j+1} = \sum_{i,\ell} a_{i,\ell}^{A_{j+1}} \mathbf{v}_{i,\ell}^{A_{j+1}}$. We have that

$$|a_{i,\ell}^{A_{j+1}}| \leq C_{13} \cdot 2^{-(K_{\mathrm{d}}-\ell)(j+1)}, \tag{35}$$

from (34). In order to complete the proof, we need to show that

$$|a_{i,\ell}^{j+1}| \leq C_{15} \cdot 2^{-(K_{\mathrm{d}}-\ell)(j+1)}, \tag{36}$$

with C_{15} independent of j. Note that the key difference between this step and the result of the previous step of the proof is that A_{j+1} is not necessarily aligned with the center of the hypercube X_{j+1}. Hence, proving (36) is more general than the result in the previous step. Using the binomial theorem, we have the following relationship between the two sets of coefficients:

$$a_{i,\ell}^{j+1} = \sum_{(i',\ell')\in\mathcal{S}(i,\ell)} a_{i',\ell'}^{A_{j+1}} \left(\prod_{k=1}^{M-1} \binom{i'_k}{i'_k - i_k} (-\eta_k)^{i'_k - i_k} \right),$$

where $\mathcal{S}(i,\ell) = \{(i',\ell') : \ell' \geq \ell; i'_1 \geq i_1, \cdots, i'_{M-1} \geq i_{M-1}; i'_1 + i'_2 + \cdots + i'_{M-1} = \ell'\}$. The outer sum is finite and $|\eta_k| \leq C_{16} \cdot 2^{-(j+1)}$ for all k, thus establishing (36):

$$\begin{aligned}
|a_{i,\ell}^{j+1}| &\leq \sum_{(i',\ell')\in\mathcal{S}(i,\ell)} |a_{i',\ell'}^{A_{j+1}}| \left(\prod_{k=1}^{M-1} C_{17} \cdot |\eta_k|^{i'_k - i_k} \right) \\
&\leq \sum_{(i',\ell')\in\mathcal{S}(i,\ell)} |a_{i',\ell'}^{A_{j+1}}| \cdot (C_{17} \cdot C_{16})^{M-1} \cdot 2^{-(j+1)\sum_{k=1}^{M-1} i'_k - i_k} \\
&\leq \sum_{(i',\ell')\in\mathcal{S}(i,\ell)} C_{13} \cdot 2^{-(K_{\mathrm{d}} - \ell')(j+1)} \cdot (C_{17} \cdot C_{16})^{M-1} \cdot 2^{-(j+1)(\ell' - \ell)} \\
&\leq C_{15} \cdot 2^{-(K_{\mathrm{d}} - \ell)(j+1)}.
\end{aligned}$$

Here, we use (35) for the third inequality. □

F Proof of Theorem 6

We begin by providing a simple performance bound (which we improve upon below) for the approximation scheme described in Sec. 4.3 applied to $f_\sharp^{\mathrm{c}}$. At scale J, the $(M-2)$-dimensional intersection manifold passes through $O(2^{(M-2)J})$ hypercubes. Let one such hypercube be denoted by $X_{J,\sharp}$. The squared-L_2 error in $X_{J,\sharp}$ due to the use of a surflet (that is ill-suited for representing intersections) can be approximated by the volume of the hypercube and is equal to $O(2^{-MJ})$. Therefore, the total squared-L_2 approximation error in representing the $(M-2)$-dimensional intersection manifold is given by $O(2^{-2J})$ (for every M). Comparing this result to the bound in (19), we see that the approximation performance achieved by our surflet-based representation scheme applied to $f_\sharp^{\mathrm{c}}$ is $\|f_\sharp^{\mathrm{c}} - \widehat{f_{\sharp,N}^{\mathrm{c}}}\|_{L_2}^2 \lesssim \left(\frac{1}{N}\right)^{\frac{2}{M-1}}$. Hence, the representation scheme from Sec. 4.3 applied to $f_\sharp^{\mathrm{c}}$ achieves optimal approximation performance for $K_{\mathrm{d}} = 2$, but sub-optimal performance for $K_{\mathrm{d}} > 2$. Thus, the smoothness threshold $K_{\mathrm{d}}^{\mathrm{th}} = 2$. This performance is still better than that achieved by wavelets, which treat the discontinuities as $\mathcal{C}^1$-smooth functions regardless of any additional structure in the discontinuities, and thus have a smoothness threshold of $K_{\mathrm{d}}^{\mathrm{th,wl}} = 1$ (i.e., the performance achieved by wavelets is $\|f_\sharp^{\mathrm{c}} - \widehat{f_{\sharp,N}^{\mathrm{c}}}\|_{L_2}^2 \lesssim \left(\frac{1}{N}\right)^{\frac{1}{M-1}}$).

Using more sophisticated analysis we now improve upon the performance bound described above to show that the approximation scheme has a smoothness threshold $K_{\mathrm{d}}^{\mathrm{th}}$ greater than two, thus increasing the range of K_{d} for which we achieve optimal performance. In order to improve the performance bound, we consider the scenario where the approximation scheme further subdivides those hypercubes at scale J containing intersections. At scale J, the scheme described in

Sec. 4.3 uses $N = O(2^{(M-1)J})$ terms to construct an approximation (analyzed in detail in Appendix C). This suggests that an additional $O(2^J)$ mini-hypercubes could be used within each of the $O(2^{(M-2)J})$ hypercubes that contain the $(M-2)$-dimensional intersection manifold (again, let one such hypercube be denoted by $X_{J,\sharp}$), while still maintaining the same order for N. Let the sidelength of the mini-hypercubes be 2^{-s}. Each of these mini-hypercubes is labeled with a surflet (chosen from a dictionary at scale s). Within each $X_{J,\sharp}$, the approximation scheme would ideally use the $O(2^J)$ mini-hypercubes to only approximate those regions that contain the intersection manifold rather than to tile the entire hypercube $X_{J,\sharp}$. Using this idea we compute the sidelength, and consequently the smoothness threshold $K_{\mathrm{d}}^{\mathrm{th}}$, as follows:

- $M > 2$**:** The number of mini-hypercubes used in each $X_{J,\sharp}$ is $O(2^J)$. We would like all these mini-hypercubes of sidelength 2^{-s} to approximate regions within $X_{J,\sharp}$ that contain the intersection manifold. This implies that $2^J \sim 2^{(M-2)(s-J)}$, which results in $s = \left\lceil \frac{(M-1)J}{M-2} \right\rceil$. The total squared-$L_2$ error due to the use of these mini-hypercubes in each $X_{J,\sharp}$ is $O(2^J \cdot 2^{-Ms}) = O(2^J \cdot 2^{\frac{-M(M-1)J}{M-2}})$. The total squared-$L_2$ error over all of the $O(2^{(M-2)J})$ $X_{J,\sharp}$'s due to the mini-hypercubes is given by $O(2^J \cdot 2^{\frac{-M(M-1)J}{M-2}} \cdot 2^{(M-2)J}) = O(2^{\frac{-2(M-1)J}{M-2}})$. Comparing with (19), we have that $K_{\mathrm{d}}^{\mathrm{th}} = \frac{2(M-1)}{M-2}$.
- $M = 2$**:** In this case, discontinuities intersect at points. Therefore, only a constant number of mini-hypercubes are needed inside each $X_{J,\sharp}$. As a result, the number of mini-hypercubes that are required to cover the intersection manifold does not grow with scale. Choosing $s = \left\lceil \frac{K_{\mathrm{d}}J}{M} \right\rceil$, we see that the total squared-L_2 error due to the use of these mini-hypercubes in each $X_{J,\sharp}$ is $O(2^{-Ms}) = O(2^{-K_{\mathrm{d}}J})$. The total squared-$L_2$ error over all of the hypercubes $X_{J,\sharp}$ (a constant number) due to the mini-hypercubes is also given by $O(2^{-K_{\mathrm{d}}J})$. Comparing with (19), we see that the scheme in Sec. 4.3 achieves optimal approximation performance for every K_{d}.

Note that the analysis of the approximation scheme as described above requires explicit information about the location of the intersections; however, an approximation scheme based on L_2-best surflets (with the dictionary containing regular and "mini"-surflets) would not require such explicit information but would still achieve the same performance. □

G Proof of Theorem 7

According to the prototype algorithm, there are three sources of error in the approximation — quantizing wavelet coefficients that are kept in the tree, pruning Type S nodes (and their descendant subtrees) from scale $\frac{J}{m}$ to scale $\frac{J}{n}$, and approximating Type D discontinuity nodes at scale $\frac{J}{n}$ (and their descendant subtrees) by surfprint atoms. The terms used in constructing the approximation include Type S wavelet coefficients, Type D wavelet coefficients, and surfprint atoms. We will analyze the approximation error and number of terms separately before calculating the approximation rate of the surfprint-based representation.

Distortion: The sources of distortion contribute in the following manner:

- *Quantizing wavelet coefficients:* There are $N_{1:\frac{J}{m}} = O(2^{\frac{MJ}{m}})$ wavelet nodes up to scale $\frac{J}{m}$. The number of Type D nodes (which are not pruned) at scales $j \in [\frac{J}{m}+1, \frac{J}{n}]$ is given by $N_{j,\mathrm{D}} = O(2^{(M-1)j})$. This implies that the total number of Type D nodes at these scales is bounded by $N_{\frac{J}{m}+1:\frac{J}{n},\mathrm{D}} = O(2^{\frac{J(M-1)}{n}}) = O(2^{\frac{MJ}{m}})$ because $\frac{m}{n} = \frac{M}{M-1}$. Using (8), each wavelet coefficient in the approximation is quantized up to resolution $(\Delta^{K_\mathrm{s}^\mathrm{wl}})^2 = 2^{-\frac{J}{m}(2K_\mathrm{s}^\mathrm{wl}+M)}$. Therefore, the total quantization distortion is $O(2^{-\frac{2K_\mathrm{s}^\mathrm{wl}J}{m}})$.

- *Pruning Type S nodes:* First, we consider Type S nodes at scale $\frac{J}{m}$. The magnitude of wavelet coefficients for Type S smooth nodes decays as $O(2^{-(K_\mathrm{s}^\mathrm{wl}+M/2)j})$ [25]. The squared-L_2 error from a single pruning at scale $\frac{J}{m}$ is given by:

$$\sum_{j=\frac{J}{m}}^{\infty} 2^{M(j-\frac{J}{m})} \cdot 2^{-j(2K_\mathrm{s}^\mathrm{wl}+M)} \lesssim 2^{-\frac{J}{m}(M+2K_\mathrm{s}^\mathrm{wl})}, \tag{37}$$

  where a Type S node at scale J/m has $2^{M(j-\frac{J}{m})}$ children nodes at scale j. There are $N_{\frac{J}{m},\mathrm{S}} = O(2^{\frac{MJ}{m}})$ Type S nodes at scale $\frac{J}{m}$. Therefore, the total distortion from pruning Type S nodes at scale $\frac{J}{m}$ is $O(2^{-\frac{2K_\mathrm{s}^\mathrm{wl}J}{m}})$. Second, we consider Type S nodes (not previously pruned) at deep scales greater than $\frac{J}{m}$. The error given by (37) also serves as an upper bound for every Type S pruning from scale $\frac{J}{m}+1$ to scale $\frac{J}{n}$. For a Type S node at these scales to have not been previously pruned, it must have a Type D parent. Because $N_{\frac{J}{m}+1:\frac{J}{n},\mathrm{D}} = O(2^{\frac{MJ}{m}})$, the total error due to pruning is $O(2^{\frac{-2K_\mathrm{s}^\mathrm{wl}J}{m}} + 2^{-\frac{J(M+2K_\mathrm{s}^\mathrm{wl})}{m}} \cdot 2^{\frac{MJ}{m}}) = O(2^{\frac{-2K_\mathrm{s}^\mathrm{wl}J}{m}})$.

- *Using surfprint approximations:* The number of Type D discontinuity nodes at scale $\frac{J}{n}$ (approximated by surfprint atoms) is $N_{\frac{J}{n},\mathrm{D}} = N_{\frac{J}{n},\mathrm{SP}} = O(2^{\frac{(M-1)J}{n}})$. The error due to each surfprint approximation is given by $O(2^{\frac{-K_\mathrm{d}^\mathrm{sp}J}{n}} \cdot 2^{\frac{-(M-1)J}{n}} + 2^{\frac{-(2K_\mathrm{s}^\mathrm{sp}+M)J}{n}})$. (This error is bounded by the squared-L_2 error of the quantized piecewise polynomial surflet approximation over each hypercube $X_{J/n}$, extended if necessary to cover the supports of the wavelets.) Therefore, the total error due to surfprint approximations is given by $O(2^{-\min(\frac{K_\mathrm{d}^\mathrm{sp}J}{n}, \frac{(2K_\mathrm{s}^\mathrm{sp}+1)J}{n})})$.

Thus, the total squared distortion is given by

$$\left\| f^\mathrm{s} - \widehat{f}_J^\mathrm{s} \right\|_{L_2}^2 = O(2^{-\min(\frac{2K_\mathrm{s}^\mathrm{wl}J}{m}, \frac{K_\mathrm{d}^\mathrm{sp}J}{n}, \frac{(2K_\mathrm{s}^\mathrm{sp}+1)J}{n})}). \tag{38}$$

Number of terms: The following three types of terms are used in assembling an approximation to $f^\mathrm{s} \in \mathcal{F}_\mathrm{S}(M, K_\mathrm{d}, K_\mathrm{s})$:

- *Coarse-scale wavelets:* The total number wavelets used at coarse scales is

$$N_{1:\frac{J}{m}} = O(2^{\frac{MJ}{m}}) = O(2^{(M-1)\frac{J}{n}}).$$

- *Intermediate Type D wavelets:* The number of Type D nodes used at intermediate scales is $N_{\frac{J}{m}+1:\frac{J}{n},\mathrm{D}} = O(2^{\frac{(M-1)J}{n}})$.

- *Surfprints:* The total number of surfprints used in the approximation is $N_{\frac{J}{n},\mathrm{SP}} = O(2^{\frac{(M-1)J}{n}})$.

Thus, the total number of terms, N, used in assembling the approximation is

$$N = N_{1:\frac{J}{m}} + N_{\frac{J}{m}+1:\frac{J}{n},\mathrm{D}} + N_{\frac{J}{n},\mathrm{SP}} = O(2^{\frac{(M-1)J}{n}}) = O(2^{\frac{MJ}{m}}). \tag{39}$$

Combining (38) and (39), we get the following approximation rate for the performance of our prototype surfprint-based approximation scheme:

$$\left\| f^{\mathrm{s}} - \widehat{f^{\mathrm{s}}_N} \right\|_{L_2}^2 \lesssim \left(\frac{1}{N}\right)^{\min(\frac{K_{\mathrm{d}}^{\mathrm{sp}}}{M-1}, \frac{2K_{\mathrm{s}}^{\mathrm{wl}}}{M}, \frac{2K_{\mathrm{s}}^{\mathrm{sp}}+1}{M-1})}.$$

The conditions on the operational smoothness orders $K_{\mathrm{d}}^{\mathrm{sp}}$, $K_{\mathrm{s}}^{\mathrm{wl}}$, and $K_{\mathrm{s}}^{\mathrm{sp}}$ (as specified in Sec. 5.3) ensure that the proper terms dominate the decay in this expression and it matches the optimal asymptotic approximation rate of Theorem 2. □

H Proof of Theorem 8

Consider an N-term approximation $\widehat{f^{\mathrm{s}}_N}$ to f^{s} constructed by the scheme in Sec. 5.5. The distortion between $\widehat{f^{\mathrm{s}}_N}$ and f^{s} is given by (38). We only need to analyze the number of bits required to encode this approximation.

Number of bits: We encode the topology of the tree and the quantized wavelet and surfprint terms.

- To encode the structure of the tree, we use $O(1)$ bits to encode each node in the tree:

$$R_1 = O(N) = O(2^{\frac{(M-1)J}{n}}) = O(2^{\frac{MJ}{m}}), \tag{40}$$

  from (39).

- The number of possible quantization bins for a wavelet coefficient $\langle f^{\mathrm{s}}, w_{X_j}\rangle$ at scale j is given by

$$\mathrm{Bins}(j) = O(2^{\frac{J}{m}(K_{\mathrm{s}}^{\mathrm{wl}}+\frac{M}{2})-j}),$$

  based on the quantization step-size (8) and the fact that a wavelet coefficient at scale j near the $(M-1)$-dimensional discontinuity decays as 2^{-j} [25]. Thus, the number of bits required to encode wavelet coefficients is given by

$$R_2 = \sum_{j=0}^{\frac{J}{m}} (N_{j,\mathrm{S}} + N_{j,\mathrm{D}}) \log(\mathrm{Bins}(j)) + \sum_{j=\frac{J}{m}+1}^{\frac{J}{n}} N_{j,\mathrm{D}} \log(\mathrm{Bins}(j)) = O(J2^{M\frac{J}{m}}). \tag{41}$$

- The number of bits required to encode surfprint coefficients at scale $\frac{J}{n}$ is

$$R_3 = N_{\frac{J}{n},\mathrm{SP}} O\left((K_{\mathrm{d}}^{\mathrm{sp}} + K_{\mathrm{s}}^{\mathrm{sp}})\frac{J}{n}\right) = O(J2^{(M-1)\frac{J}{n}}). \tag{42}$$

Combining (38), (40), (41), and (42), we obtain the desired result. □

I Proof of Theorem 9

Let X_J be an M-dimensional dyadic hypercube[12] with $J \leq J_{\text{vox}} := \frac{\min(\pi_i)}{K_{\text{d}}}$. Let $\widehat{f_L^{\text{c}}}$ be the continuous L_2-best surflet fit to f^{c} over X_J. We know from (4) that

$$\left\| f^{\text{c}} - \widehat{f_L^{\text{c}}} \right\|_{L_2(X_J)}^2 = O(2^{-J(K_{\text{d}}+M-1)}). \tag{43}$$

We assume that the values of the discretized function $\widetilde{f_\pi^{\text{c}}}$ are obtained by averaging f^{c} over each voxel. The distortion between f^{c} and $\widetilde{f_\pi^{\text{c}}}$ over X_J is nonzero only over voxels through which the discontinuity b^{c} passes. The squared-L_2 distortion over each such voxel is $O(2^{-(\pi_1+\cdots+\pi_M)})$. Also, the number of voxels through which b^{c} passes within $X_J = O\left(2^{\sum_{i=1}^{M-1}(\pi_i - J)} \left\lceil \left(\Omega \cdot 2^{-\min(\pi_i)_{i=1}^{M-1}}\right) / \left(2^{-\pi_M}\right) \right\rceil\right)$, where Ω is the universal derivative bound (Sec. 2.1). Thus, we have

$$\left\| f^{\text{c}} - \widetilde{f_\pi^{\text{c}}} \right\|_{L_2(X_J)}^2 = O(2^{-J(M-1)-\min(\pi_i)}) = O(2^{-J(K_{\text{d}}+M-1)}), \tag{44}$$

where the second equality is due to fact that $J \leq \frac{\min(\pi_i)}{K_{\text{d}}}$. Note that we define $\widetilde{f_\pi^{\text{c}}}$ as a continuous function (constant over each M-dimensional voxel) in order to compare with f^{c}. Similarly, one can check that

$$\left\| \widehat{f_L^{\text{c}}} - \widetilde{\widehat{f_{L,\pi}^{\text{c}}}} \right\|_{L_2(X_J)}^2 = O(2^{-J(K_{\text{d}}+M-1)}), \tag{45}$$

where $\widetilde{\widehat{f_{L,\pi}^{\text{c}}}}$ is the sampled version of $\widehat{f_L^{\text{c}}}$.

Equations (44) and (45) indicate that at scale J, voxelization effects are comparable to the approximations afforded by surflets. Essentially, then, all of the approximation results for surflets at this scale are preserved when applied to the voxelized function. In particular, combining (43), (44), and (45), we have the following result:

$$\begin{aligned} \left\| \widetilde{f_\pi^{\text{c}}} - \widetilde{\widehat{f_{L,\pi}^{\text{c}}}} \right\|_{L_2(X_J)}^2 &\leq \left(\left\| f^{\text{c}} - \widetilde{f_\pi^{\text{c}}} \right\|_{L_2(X_J)} + \left\| f^{\text{c}} - \widehat{f_L^{\text{c}}} \right\|_{L_2(X_J)} + \left\| \widehat{f_L^{\text{c}}} - \widetilde{\widehat{f_{L,\pi}^{\text{c}}}} \right\|_{L_2(X_J)} \right)^2 \\ &= O(2^{-J(K_{\text{d}}+M-1)}). \end{aligned}$$

Thus, discrete surflets are as effective on the discrete block as continuous surflets are on the corresponding continuous block (see Appendix C). However, we are only provided with the discrete function $\widetilde{f_\pi^{\text{c}}}$ and would like to use ℓ_2-best surflets on dyadic blocks of $\widetilde{f_\pi^{\text{c}}}$. Let $\widehat{\widetilde{f_\pi^{\text{c}}}}$ denote the discrete ℓ_2-best surflet fit to $\widetilde{f_\pi^{\text{c}}}$ over X_J.

By definition, $\|\widetilde{f_\pi^{\text{c}}} - \widehat{\widetilde{f_\pi^{\text{c}}}}\|_{L_2(X_J)}^2 \leq \|\widetilde{f_\pi^{\text{c}}} - \widetilde{\widehat{f_{L,\pi}^{\text{c}}}}\|_{L_2(X_J)}^2$. Thus, we have that

$$\left\| \widetilde{f_\pi^{\text{c}}} - \widehat{\widetilde{f_\pi^{\text{c}}}} \right\|_{L_2(X_J)}^2 = O(2^{-J(K_{\text{d}}+M-1)}). \tag{46}$$

It remains to be shown that ℓ_2-best surflets can be predicted across scales. The proof of this fact is analogous to the proof of Theorem 5. □

[12] We omit an additive constant that may be added to J_{vox} to ensure a more exact agreement with the voxelization breakdown rate.

J Proof of Corollary 10

Combining (44) and (46) from Appendix I, we have the following result:

$$\begin{aligned}\left\|f^{\mathrm{c}}-\widehat{\widetilde{f_{\pi}^{\mathrm{c}}}}\right\|_{L_2(X_J)}^2 &\leq \left(\left\|f^{\mathrm{c}}-\widetilde{f_{\pi}^{\mathrm{c}}}\right\|_{L_2(X_J)}+\left\|\widetilde{f_{\pi}^{\mathrm{c}}}-\widehat{\widetilde{f_{\pi}^{\mathrm{c}}}}\right\|_{L_2(X_J)}\right)^2\\ &= O(2^{-J(K_{\mathrm{d}}+M-1)}).\end{aligned}$$

Hence, the quality of the approximation provided by the ℓ_2-best surflets $\widehat{\widetilde{f_{\pi}^{\mathrm{c}}}}$ operating on $\widetilde{f_{\pi}^{\mathrm{c}}}$ to the continuous function f^{c}, with $\widehat{\widetilde{f_{\pi}^{\mathrm{c}}}}$ considered to be continuous (constant over each M-dimensional voxel), is as good as the approximation performance provided by continuous L_2-best surflet fits to f^{c} (see Appendix C). The rest of the proof follows in an analogous manner to the proof of Theorem 5. □

K Proof of Theorem 11

In smooth regions of $\widetilde{f_{\pi}^{\mathrm{s}}}$, discrete wavelet coefficients (corresponding to a discrete wavelet basis with $K_{\mathrm{s}}^{\mathrm{wl}}$ vanishing moments applied to $\widetilde{f_{\pi}^{\mathrm{s}}}$) decay as $O(2^{-(K_{\mathrm{s}}^{\mathrm{wl}}+M/2)j})$ [25]. This is the same decay rate that continuous wavelet coefficients obey on smooth regions of f^{s} (see Appendix G). Therefore, the total error due to the use of discrete wavelets to approximate $\widetilde{f_{\pi}^{\mathrm{s}}}$, the quantization of the corresponding discrete wavelet coefficients, and the pruning of Type S nodes is analogous to the corresponding error in the continuous case analyzed in Appendix G.

What remains to be analyzed is the distortion due to the use of discrete surfprint approximations. This analysis is analogous to the analysis in Appendix I. Consider a Type D node at scale $\frac{J}{n}$ to which a surfprint approximation is assigned. Let the hypercube corresponding to this node be $X_{\frac{J}{n}}$. First, we have the following distortion between f^{s} and the L_2-best surfprint fit to f^{s} over $X_{\frac{J}{n}}$, $\widehat{f_{\mathrm{sp},L}^{\mathrm{s}}}$ (see Appendix G):

$$\begin{aligned}\left\|f^{\mathrm{s}}-\widehat{f_{\mathrm{sp},L}^{\mathrm{s}}}\right\|_{L_2(X_{\frac{J}{n}})}^2 &= O\left(2^{\frac{-(K_{\mathrm{d}}^{\mathrm{sp}}+M-1)J}{n}}+2^{\frac{-(2K_{\mathrm{s}}^{\mathrm{sp}}+M)J}{n}}\right)\\ &= O\left(2^{-\frac{(M-1)J}{n}}\cdot 2^{-\min(K_{\mathrm{d}}^{\mathrm{sp}},2K_{\mathrm{s}}^{\mathrm{sp}}+1)\frac{J}{n}}\right). \qquad (47)\end{aligned}$$

Second, we characterize the error due to voxelization of f^{s} over $X_{\frac{J}{n}}$. For each voxel in $X_{\frac{J}{n}}$ through which the discontinuity passes, the squared-L_2 error between f^{s} and $\widetilde{f_{\pi}^{\mathrm{s}}}$ is given by $O(2^{-(\pi_1+\cdots+\pi_M)})$. The number of such voxels in $X_{\frac{J}{n}} = O\left(2^{\sum_{i=1}^{M-1}(\pi_i-\frac{J}{n})}\left\lceil\left(\Omega\cdot 2^{-\min(\pi_i)_{i=1}^{M-1}}\right)/\left(2^{-\pi_M}\right)\right\rceil\right)$, where Ω is the universal derivative bound (Sec. 2.1). Discretization over the smooth regions of f^{s} can be viewed as a construction of 0'th order Taylor approximations (constants) locally over each voxel. Therefore, the squared-L_2 error between f^{s} and $\widetilde{f_{\pi}^{\mathrm{s}}}$ over a voxel in the smooth region is $O\left(2^{-2\min(\pi_i)}\cdot 2^{-(\pi_1+\cdots+\pi_M)}\right)$. The number of such voxels in $X_{\frac{J}{n}} = O\left(2^{\sum_{i=1}^{M}(\pi_i-\frac{J}{n})}\right)$. Thus, the

total error between f^{s} and $\widetilde{f^{\mathrm{s}}_{\pi}}$ is

$$\begin{aligned}\left\|f^{\mathrm{s}}-\widetilde{f^{\mathrm{s}}_{\pi}}\right\|^2_{L_2(X_{\frac{J}{n}})} &= O\left(2^{-\frac{(M-1)J}{n}-\min(\pi_i)}+2^{-\frac{MJ}{n}-2\min(\pi_i)}\right)\\ &= O\left(2^{-\frac{(M-1)J}{n}-\min(\pi_i)}\right)\\ &= O\left(2^{-\frac{(M-1)J}{n}}\cdot 2^{-\min(K^{\mathrm{sp}}_{\mathrm{d}},2K^{\mathrm{sp}}_{\mathrm{s}}+1)\frac{J}{n}}\right). \qquad (48)\end{aligned}$$

The last equality is due to the assumption that $\frac{J}{n}\leq\frac{\min\{\pi_i\}}{\min\{K^{\mathrm{sp}}_{\mathrm{d}},2K^{\mathrm{sp}}_{\mathrm{s}}+1\}}$. Similarly, the squared-L_2 error between a continuous surfprint atom and its discrete analog at scale $\frac{J}{n}$ is also given by

$$\left\|\widehat{f^{\mathrm{s}}_{\mathrm{sp},L}}-\widehat{\widetilde{f^{\mathrm{s}}_{\mathrm{sp},L,\pi}}}\right\|^2_{L_2(X_{\frac{J}{n}})}=O\left(2^{-\frac{(M-1)J}{n}}\cdot 2^{-\min(K^{\mathrm{sp}}_{\mathrm{d}},2K^{\mathrm{sp}}_{\mathrm{s}}+1)\frac{J}{n}}\right). \qquad (49)$$

Combining (47), (48), and (49), we have the following result:

$$\begin{aligned}\left\|\widetilde{f^{\mathrm{s}}_{\pi}}-\widehat{\widetilde{f^{\mathrm{s}}_{\mathrm{sp},L,\pi}}}\right\|^2_{L_2(X_{\frac{J}{n}})} &\leq \left(\left\|f^{\mathrm{s}}-\widetilde{f^{\mathrm{s}}_{\pi}}\right\|_{L_2(X_{\frac{J}{n}})}+\left\|f^{\mathrm{s}}-\widehat{f^{\mathrm{s}}_{\mathrm{sp},L}}\right\|_{L_2(X_{\frac{J}{n}})}+\left\|\widehat{f^{\mathrm{s}}_{\mathrm{sp},L}}-\widehat{\widetilde{f^{\mathrm{s}}_{\mathrm{sp},L,\pi}}}\right\|_{L_2(X_{\frac{J}{n}})}\right)^2\\ &= O\left(2^{-\frac{(M-1)J}{n}}\cdot 2^{-\min(K^{\mathrm{sp}}_{\mathrm{d}},2K^{\mathrm{sp}}_{\mathrm{s}}+1)\frac{J}{n}}\right). \qquad (50)\end{aligned}$$

There are $O\left(2^{\frac{(M-1)J}{n}}\right)$ Type D nodes at scale $\frac{J}{n}$. From (50), we have that the total error due to discrete surfprint approximations is given by $O\left(2^{-\min(K^{\mathrm{sp}}_{\mathrm{d}},2K^{\mathrm{sp}}_{\mathrm{s}}+1)\frac{J}{n}}\right)$. Similar to the argument made in Appendix I, we have by definition that the discrete surfprint approximation $\widehat{\widetilde{f^{\mathrm{s}}_{\mathrm{sp},\pi}}}$, constructed using the discrete ℓ_2-best surfprint atoms, satisfies $\|\widetilde{f^{\mathrm{s}}_{\pi}}-\widehat{\widetilde{f^{\mathrm{s}}_{\mathrm{sp},\pi}}}\|^2_{L_2(X_J)}\leq\|\widetilde{f^{\mathrm{s}}_{\pi}}-\widehat{\widetilde{f^{\mathrm{s}}_{\mathrm{sp},L,\pi}}}\|^2_{L_2(X_J)}$. Thus,

$$\left\|\widetilde{f^{\mathrm{s}}_{\pi}}-\widehat{\widetilde{f^{\mathrm{s}}_{\mathrm{sp},\pi}}}\right\|^2_{L_2(X_{\frac{J}{n}})}=O\left(2^{-\frac{(M-1)J}{n}}\cdot 2^{-\min(K^{\mathrm{sp}}_{\mathrm{d}},2K^{\mathrm{sp}}_{\mathrm{s}}+1)\frac{J}{n}}\right).$$

Combining this result with the arguments made in the first paragraph of this proof, the squared-L_2 error between the discrete function $\widetilde{f^{\mathrm{s}}_{\pi}}$ and the discrete wavelet/surfprint approximation $\widehat{\widetilde{f^{\mathrm{s}}_{\pi}}}$ (this approximation now represents a composite wavelet/surfprint approximation; the previous paragraph only analyzed surfprint approximations) is

$$\left\|\widetilde{f^{\mathrm{s}}_{\pi}}-\widehat{\widetilde{f^{\mathrm{s}}_{\pi}}}\right\|^2_{L_2}=O\left(2^{-\min(\frac{2K^{\mathrm{wl}}_{\mathrm{s}}J}{m},\frac{K^{\mathrm{sp}}_{\mathrm{d}}J}{n},\frac{(2K^{\mathrm{sp}}_{\mathrm{s}}+1)J}{n})}\right). \qquad (51)$$

The rest of this proof is analogous to the proof of Theorem 8 in Appendix H. □

L Proof of Corollary 12

This proof is similar to the proof in Appendix J. We begin by extending the bound provided by (48). This bound holds for Type D hypercubes at scale $\frac{J}{n}$. The total number of such Type D hypercubes

at scale $\frac{J}{n}$ is given by $O(2^{(M-1)\frac{J}{n}})$. Following the logic preceding (48), we have that the squared-L_2 error between f^{s} and $\widetilde{f^{\mathrm{s}}_{\pi}}$ over a Type S hypercube at scale j is given by $O(2^{-Mj-2\min(\pi_i)})$, and the total number of such Type S hypercubes at scale j is given by $O(2^{Mj})$. Combining these arguments with (51) from Appendix K, we have the following result:

$$\begin{aligned}\left\| f^{\mathrm{s}} - \widehat{\widetilde{f^{\mathrm{s}}_{\pi}}} \right\|_{L_2}^2 &\leq \left(\left\| f^{\mathrm{s}} - \widetilde{f^{\mathrm{s}}_{\pi}} \right\|_{L_2} + \left\| \widetilde{f^{\mathrm{s}}_{\pi}} - \widehat{\widetilde{f^{\mathrm{s}}_{\pi}}} \right\|_{L_2} \right)^2 \\ &= O\left(2^{-\min(\frac{2K_{\mathrm{s}}^{\mathrm{wl}} J}{m}, \frac{K_{\mathrm{d}}^{\mathrm{sp}} J}{n}, \frac{(2K_{\mathrm{s}}^{\mathrm{sp}}+1)J}{n})} \right).\end{aligned}$$

The rest of the proof follows in an analogous manner to the proof of Theorem 8. □

References

[1] V. Chandrasekaran, M. B. Wakin, D. Baron, and R. Baraniuk, "Compression of higher dimensional functions containing smooth discontinuities," in *Conf. on Information Sciences and Systems — CISS 2004*, Princeton, New Jersey, Mar. 2004.

[2] ——, "Surflets: A sparse representation for multidimensional functions containing smooth discontinuities," in *2004 IEEE International Symposium on Information Theory (ISIT2004)*, Chicago, Illinois, June 2004.

[3] R. A. DeVore, B. Jawerth, and B. J. Lucier, "Image compression through wavelet transform coding," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 719–746, Mar. 1992.

[4] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.

[5] Z. Xiong, K. Ramchandran, and M. T. Orchard, "Space-frequency quantization for wavelet image coding," *IEEE Trans. Image Processing*, vol. 6, no. 5, pp. 677–693, 1997.

[6] A. Cohen, I. Daubechies, O. G. Guleryuz, and M. T. Orchard, "On the importance of combining wavelet-based nonlinear approximation with coding strategies," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1895–1921, July 2002.

[7] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[8] ——, "Wedgelets: Nearly-minimax estimation of edges," *Annals of Stat.*, vol. 27, pp. 859–897, 1999.

[9] H. Krim, D. Tucker, S. Mallat, and D. Donoho, "On denoising and best signal representation," *IEEE Trans. Inform. Theory*, vol. 45, no. 7, pp. 2225–2238, 1999.

[10] R. Willett and R. Nowak, "Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, Mar. 2003.

[11] E. Mammen and A. B. Tsybakov, "Smooth discrimination analysis," *Annals of Stat.*, vol. 27, pp. 1808–1829, 1999.

[12] A. B. Tsybakov, "Optimal aggregation of classifiers in statistical learning," *Annals of Stat.*, vol. 32, no. 1, pp. 135–166, 2004.

[13] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inform. Theory*, vol. 36, no. 5, pp. 961–1005, 1990.

[14] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore, "Tree approximation and optimal encoding," *J. Appl. Comp. Harm. Anal.*, vol. 11, pp. 192–226, 2001.

[15] J. Kovacevic and M. Vetterli, "Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for R^n," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, part 2, pp. 533–555, 1992.

[16] E. J. Candès and D. L. Donoho, "Curvelets — A suprisingly effective nonadaptive representation for objects with edges," in *Curve and Surface Fitting*, A. Cohen, C. Rabut, and L. L. Schumaker, Eds. Vanderbilt University Press, 1999.

[17] M. Levoy and P. Hanrahan, "Light field rendering," in *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 1996, pp. 31–42.

[18] J. VanderWerf, N. Kourjanskaia, S. Shladover, H. Krishnan, and M. Miller, "Modeling the effects of driver control assistance systems on traffic," in *National Research Council Transportation Research Board 80th Annual Meeting*, January 2001.

[19] A. N. Kolmogorov and V. M. Tihomirov, "ϵ-entropy and ϵ-capacity of sets in functional spaces," *Amer. Math. Soc. Transl. (Ser. 2)*, vol. 17, pp. 277–364, 1961.

[20] G. F. Clements, "Entropies of several sets of real valued functions," *Pacific J. Math.*, vol. 13, pp. 1085–1095, 1963.

[21] M. A. L. Cooke, "Richard Surflet, translator and practitioner in physic," *Medical History*, vol. 25, pp. 41–56, 1981.

[22] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. San Diego: Academic Press, 1999.

[23] M. B. Wakin, J. K. Romberg, H. Choi, and R. G. Baraniuk, "Wavelet-domain approximation and compression of piecewise smooth images," *IEEE Trans. Image Processing*, May 2006.

[24] A. Lisowska, "Effective coding of images with the use of geometrical wavelets," in *Proc. Decision Support Systems Conference*, Zakopane, Poland, 2003, in Polish.

[25] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Information Theory*, vol. 44, no. 6, pp. 2435–2476, 1998.

[26] M. N. Do, P. L. Dragotti, R. Shukla, and M. Vetterli, "On the compression of two-dimensional piecewise smooth functions," in *IEEE Int. Conf. on Image Proc. — ICIP '01*, Thessaloniki, Greece, Oct. 2001.

[27] V. Chandrasekaran, M. Wakin, D. Baron, and R. Baraniuk, "Compressing Piecewise Smooth Multidimensional Functions Using Surflets: Rate-Distortion Analysis," Rice University ECE Department, Houston, TX, Tech. Rep., March 2004.

[28] J. Romberg, M. Wakin, and R. Baraniuk, "Multiscale Geometric Image Processing," in *SPIE Visual Comm. and Image Proc.*, Lugano, Switzerland, July 2003.

[29] J. K. Romberg, M. B. Wakin, and R. G. Baraniuk, "Multiscale wedgelet image analysis: Fast decompositions and modeling," in *IEEE Int. Conf. on Image Proc. — ICIP '02*, Rochester, New York, 2002.

[30] M. B. Wakin, D. L. Donoho, H. Choi, and R. G. Baraniuk, "The multiscale structure of non-differentiable image manifolds," in *Proc. Wavelets XI at SPIE Optics and Photonics*, San Diego, August 2005.

[31] F. M. J. Willems, Y. M. Shtarkov, and T. J. Tjalkens, "The context-tree weighting method: Basic properties," *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 653–664, 1995.

[32] E. Kreyszig, *Introduction to Functional Analysis with Applications*. New York: John Wiley and Sons, 1978.