Pyramid transform of manifold data via subdivision operators

Wael Mattar* and Nir Sharon
School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel
*Corresponding author: waelmattar.95@gmail.com

Multiscale transforms have become a key ingredient in many data processing tasks. With technological development we observe a growing demand for methods to cope with nonlinear data structures such as manifold values. In this paper we propose a multiscale approach for analyzing manifold-valued data using a pyramid transform. The transform uses a unique class of downsampling operators that enable a noninterpolating subdivision schemes as upsampling operators. We describe this construction in detail and present its analytical properties, including stability and coefficient decay. Next, we numerically demonstrate the results and show the application of our method to denoising and anomaly detection.

Keywords: multiscale approximation; manifold-valued curves; subdivision schemes; nonlinear operators; geometric data.

1. Introduction

Many modern applications use manifold values as a primary tool to model data, e.g., Rahman *et al.* (2005); Lunga *et al.* (2014); Frank & Ourmazd (2016). Manifolds express a global nonlinear structure with constrained, high-dimensional elements. The employment of manifolds as data models raises the demand for computational methods to address fundamental tasks, like integration, interpolation and regression, which become challenging under the manifold setting, see, e.g., Blanes & Casas (2017); Barbaresco & Gay-Balmaz (2020); Iserles *et al.* (2000); Zeilmann *et al.* (2020). We focus on constructing a multiscale representation for manifold values using a fast pyramid transform.

Multiscale transforms are standard tools in signal and image processing that enable a hierarchical analysis of an object mathematically. Customarily, the first scale in the transform corresponds to a coarse representation, and as scales increase, so do the levels of approximation (Mallat, 1999). The pyramid transform uses a refinement or upsampling operator together with a corresponding subsampling operator for the construction of a fast multiscale representation of signals (Donoho, 1992; Wallner, 2020). The simplicity of this powerful method opened the door for many applications. Naturally, recent years found generalizations of multiscale representations for manifold values as well as manifold-valued pyramid transforms (Storath & Weinmann, 2020). Contrary to the classical, linear settings, where upsampling operators are often linear and global, e.g., polynomial interpolation, refinement operators to manifolds values are mostly nonlinear and local operators. One such class of operators arises in subdivision schemes.

Subdivision schemes are powerful yet computationally efficient tools for producing smooth objects from discrete sets of points. These schemes are defined by repeatedly applying a subdivision operator that refines discrete sets. The subdivision refinements, which serve as upsampling operators, give rise to a natural connection between multiscale representations and subdivision schemes (Daubechies, 1992, Chapter 6). In recent years subdivision operators were adapted to manifold data and nonlinear

geometries by various methods, and so have been their induced multiscale transforms, see Wallner (2020) for an overview.

Multiscale transforms, based upon subdivision operators, commonly use interpolating subdivision schemes, i.e., operations that preserve the coarse objects through refinements. This standard facilitates the calculation of the missing detail coefficients at all scales. Particularly, coefficients associated with the interpolating values are systematically zeroed, and do not have to be saved or processed in the following analysis levels. Therefore, this property of interpolating multiscale transforms makes the crux of many state-of-the-art algorithms, including data compression, see, e.g., Lv *et al.* (2018). On the other hand, the notion of noninterpolating pyramid transforms did not receive equal attention despite the popularly used noninterpolating subdivision schemes. A classic example of a widespread family of noninterpolating subdivision operators is the well-known B-spline, see De Boor (1978); Lane & Riesenfeld (1980).

The main challenge behind constructing a noninterpolating pyramid transform revolves around the question of calculating the multiscale details. In particular, given a noninterpolating subdivision scheme, the corresponding subsampling operators involve applying infinitely-supported real-valued sequences. Therefore, care must be taken when realizing and implementing these operators. In this paper, we introduce a novel family of pyramid transforms suitable for noninterpolating subdivision schemes. Our multiscale transforms decompose manifold-valued sequences in a similar pyramidical fashion to the interpolating ones. Specifically, the noninterpolating transforms' construction relies on the recently introduced decimation operators, see Dyn & Zhuang (2020), which are employed as subsampling operators. From the interpolating point of view, the decimation operation coincides with the simple downsampling operation, taking all even-indexed elements.

Our contribution in this paper also covers the computability of the linear decimation operators and the linear operators' adaptation to cope with manifold data. In particular, it is not possible to implement decimation operators since they involve infinitely supported sequences. Therefore, we approximate these operators with affine averages of finite elements, which successfully lead to the desired mathematical results. We also derive an analytic condition for decimation operators termed 'decimation-safety', and show that all our operators over manifold data satisfy the condition. We prove that multiscale transforms associated with decimation-safe operators, together with their corresponding inverses, enjoy coefficient decay and stability properties. The outcomes are an essential feature of the multiscale transform of manifold data, and can significantly contribute to various applications and scientific questions.

We conclude the paper with several numerical demonstrations of both the theoretical results we obtained and applications of data processing. Specifically, we provide examples of the use of our method for denoising and anomaly detection on synthetically generated manifold data. All figures and examples were generated using a code package that complements the paper and is available online for reproducibility.

2. Preliminaries

2.1 *Linear univariate subdivision schemes*

In the functional setting a linear binary subdivision scheme $\mathcal{S}$ operates on a real-valued bi-infinite sequence $\boldsymbol{c} = \{c_k \in \mathbb{R} \mid k \in \mathbb{Z}\}$. Applying the subdivision scheme on $\boldsymbol{c}$ yields a sequence $\mathcal{S}(\boldsymbol{c})$, which is associated with the values over the refined grid $2^{-1}\mathbb{Z}$. This process is repeated infinitely and results in values defined on the dyadic rationals, which form a dense set over the real line. If the generated values

of the repeated process for any sequence c converge uniformly at the dyadic points to the values of a continuous function, we term the subdivision scheme convergent and treat the function as its limit, see, e.g., Dyn (2002a). We denote a linear, binary refinement rule of a univariate subdivision scheme $\mathcal{S}$ with a finitely supported mask $\boldsymbol{\alpha}$ by

$$\mathcal{S}_{\boldsymbol{\alpha}}(\boldsymbol{c})_k = \sum_{i \in \mathbb{Z}} \alpha_{k-2i} c_i, \quad k \in \mathbb{Z}. \tag{1}$$

Depending on the parity of the index k, the refinement rule (1) can be split into two rules. Namely,

$$\mathcal{S}_{\boldsymbol{\alpha}}(\boldsymbol{c})_{2k} = \sum_{i \in \mathbb{Z}} \alpha_{2i} c_{k-i} \quad \text{and} \quad \mathcal{S}_{\boldsymbol{\alpha}}(\boldsymbol{c})_{2k+1} = \sum_{i \in \mathbb{Z}} \alpha_{2i+1} c_{k-i}, \quad k \in \mathbb{Z}. \tag{2}$$

For more details we encourage the reader to see Dyn (2002b). Moreover, the scheme (1) can be written as the convolution $\mathcal{S}_{\boldsymbol{\alpha}}(\boldsymbol{c}) = \boldsymbol{\alpha} * (\boldsymbol{c} \uparrow 2)$, where

$$(\boldsymbol{c} \uparrow 2)_k = \begin{cases} c_{k/2}, & k \text{ is even,} \\ 0, & \text{otherwise,} \end{cases} \quad k \in \mathbb{Z},$$

is the *upsampled sequence* $\boldsymbol{c}$.

A subdivision scheme is termed interpolating if $\mathcal{S}_{\boldsymbol{\alpha}}(\boldsymbol{c})_{2k} = c_k$ for all $k \in \mathbb{Z}$. Equivalently, $\mathcal{S}_{\boldsymbol{\alpha}}(\boldsymbol{c}) = \boldsymbol{c} \uparrow 2$ over the even indices. A necessary condition for the convergence of a subdivision scheme with the refinement rule (1), see e.g., Dyn (1992), is

$$\sum_{i \in \mathbb{Z}} \alpha_{2i} = \sum_{i \in \mathbb{Z}} \alpha_{2i+1} = 1. \tag{3}$$

Henceforth, we assume that any subdivision operator mentioned is of convergent subdivision schemes. Moreover, we refer to the masks that satisfy (3) as *shift invariant*. The reason being is that applying a subdivision scheme with shift invariant mask on a shifted data points results with precisely the shifted original outcome. Note that an invariant rule (2) is a weighted average that can be interpreted as the *center of mass* of elements c_{k-i}, with the components of $\boldsymbol{\alpha}$ as their weights. This interpretation is fundamental for the adaptation of linear subdivision schemes to manifold data, as we present next.

2.2 *The Riemannian analogue of a linear subdivision scheme*

Subdivision schemes with shift invariant masks were adapted to manifold-valued data via different methods and approaches, see e.g., Rahman *et al.* (2005); Wallner & Dyn (2005); Grohs (2010a); Dyn & Sharon (2017b). One natural extension of a linear subdivision scheme (1) to manifold-valued data can be done with the help of the Riemannian structure. Let $\mathcal{M}$ be a Riemannian manifold equipped with Riemannian metric, which we denote by $\langle \cdot, \cdot \rangle$. The Riemannian geodesic distance $\rho(\cdot, \cdot) \colon \mathcal{M}^2 \to \mathbb{R}^+$ is

$$\rho(x, y) = \inf_{\Gamma} \int_a^b |\dot{\Gamma}(t)| \, \mathrm{d}t, \tag{4}$$

where $\Gamma \colon [a, b] \to \mathcal{M}$ is a curve connecting points $\Gamma(a) = x$ and $\Gamma(b) = y$, and $|\cdot|^2 = \langle \cdot, \cdot \rangle$.

The linear subdivision scheme (1) associated with an invariant mask (3) can be characterized as the unique solution of the optimization problem

$$\mathcal{S}_{\boldsymbol{\alpha}}(\boldsymbol{c})_k = \arg\min_{x \in \mathbb{R}} \sum_{i \in \mathbb{Z}} \alpha_{k-2i} \|x - c_i\|^2, \quad k \in \mathbb{Z}, \tag{5}$$

where $\| \cdot \|$ denotes the standard Euclidean norm. This is an alternative formulation for (1) as the Euclidean center of mass.

For an $\mathcal{M}$-valued sequence $\boldsymbol{c}$ we transfer the optimization problem (5) to $\mathcal{M}$ by replacing the Euclidean distance with the Riemannian geodesic distance (4). We denote by $\mathcal{T}_{\boldsymbol{\alpha}}$ the Riemannian analogue of the linear subdivision scheme $\mathcal{S}_{\boldsymbol{\alpha}}$; given a mask $\boldsymbol{\alpha}$ we define the adapted subdivision scheme as

$$\mathcal{T}_{\boldsymbol{\alpha}}(\boldsymbol{c})_k = \arg\min_{x \in \mathcal{M}} \sum_{i \in \mathbb{Z}} \alpha_{k-2i} \rho(x, c_i)^2, \quad k \in \mathbb{Z}. \tag{6}$$

When the solution of (6) exists uniquely, we term the solution as the *Riemannian center of mass* (Grove & Karcher, 1973). It is also termed Karcher mean for matrices and Frèchet mean in more general metric spaces, see Karcher (2014).

The global well-definedness of (6) when $\alpha_k \geq 0$ is studied in Kobayashi & Nomizu (1963). Moreover, in the framework where $\mathcal{M}$ has a nonpositive sectional curvature, if the mask $\boldsymbol{\alpha}$ is shift invariant, then a globally unique solution for problem (6) can be found, see e.g., Karcher (1977); Hardering (2015); Sander (2016). Recent studies of manifolds with positive sectional curvature show necessary conditions for uniqueness on the spread of points with respect to the injectivity radius of $\mathcal{M}$ (Dyer *et al.*, 2016; Hüning & Wallner, 2020). We focus our attention on $\mathcal{M}$-valued sequences $\boldsymbol{c}$ that are admissible in the sense that $\mathcal{T}_{\boldsymbol{\alpha}}(\boldsymbol{c})$ is uniquely defined for any shift invariant mask $\boldsymbol{\alpha}$, i.e., problems (6) have unique solutions.

We interpret many alternative methods for adapting subdivision operators to manifolds as finite approximations for the Riemannian center of mass (6). This includes, for example, the exp-log methods (Rahman *et al.*, 2005; Grohs & Wallner, 2012), repeated binary averaging (Wallner & Dyn, 2005; Dyn & Sharon, 2017a) and inductive means (Dyn & Sharon, 2017b).

2.3 *Interpolating linear multiscale transform*

The notion of pyramid transforms is to represent a high-resolution sequence of data points as a pyramid consisting of a coarse approximation in addition to the multiscale layers, each corresponding to a different scale, see, e.g., Grohs (2010b); Dyn & Zhuang (2020). In this section we briefly review an interpolating multiscale transform, see, e.g., Donoho (1992); Harten (1996).

With the help of an interpolating subdivision scheme $\mathcal{S}_{\boldsymbol{\alpha}}$ a high resolution real-valued sequence $\boldsymbol{c}^{(1)}$ associated with the values over the grid $2^{-1}\mathbb{Z}$ can be decomposed into a coarse (low resolution) sequence $\boldsymbol{c}^{(0)}$ over the integers, together with a sequence of detail coefficients $\boldsymbol{d}^{(1)}$ over the grid $2^{-1}\mathbb{Z}$ by letting

$$\boldsymbol{c}^{(0)} = \boldsymbol{c}^{(1)} \downarrow 2 \quad \text{and} \quad \boldsymbol{d}^{(1)} = \boldsymbol{c}^{(1)} - \mathcal{S}_{\boldsymbol{\alpha}} \boldsymbol{c}^{(0)}, \tag{7}$$
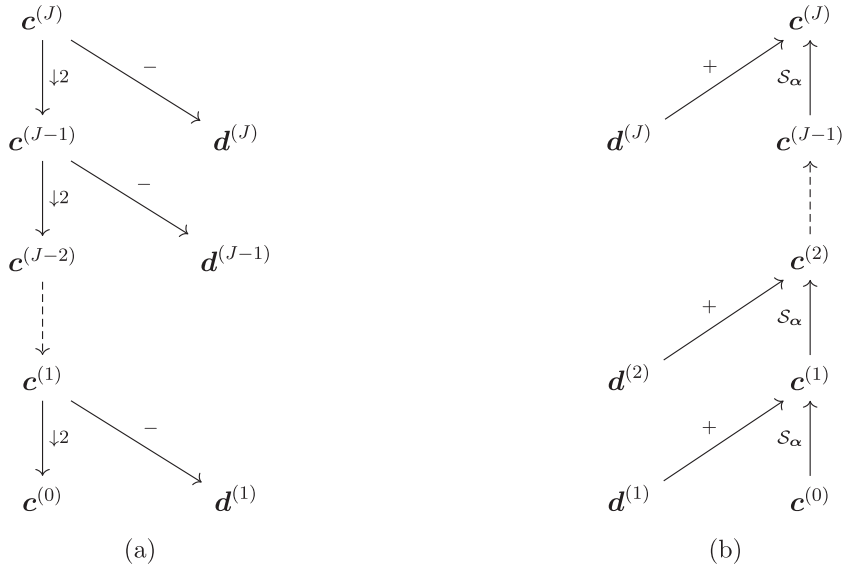
FIG. 1. The pyramid transform. On the left, the analysis (8), on the right, the synthesis (9).

where $\downarrow 2$ is the *downsampling* operator given by $(c \downarrow 2)_k = c_{2k}$ for all $k \in \mathbb{Z}$. In the same manner of decomposition (7), given a real-valued sequence $c^{(J)}$, $J \in \mathbb{N}$ associated with the values over the fine grid $2^{-J}\mathbb{Z}$, it can be recursively decomposed by

$$c^{(\ell-1)} = c^{(\ell)} \downarrow 2, \quad d^{(\ell)} = c^{(\ell)} - \mathcal{S}_\alpha c^{(\ell-1)}, \quad \ell = 1, 2, \ldots, J. \tag{8}$$

The process (8) yields a pyramid of sequences $\{c^{(0)}; d^{(1)}, \ldots, d^{(J)}\}$, where $c^{(0)}$ is the coarse approximation coefficients given over the integers, and $d^{(\ell)}$, $\ell = 1, 2, \ldots, J$ are the detail coefficients at level ℓ, given over the values of the grids $2^{-\ell}\mathbb{Z}$. We obtain synthesis by the following iterations,

$$c^{(\ell)} = \mathcal{S}_\alpha c^{(\ell-1)} + d^{(\ell)}, \quad \ell = 1, 2, \ldots, J, \tag{9}$$

which is the inverse transform of (8). At index $k \in \mathbb{Z}$ the detail coefficient $d_k^{(\ell)}$ measures the agreement between $c_k^{(\ell)}$ and $(\mathcal{S}_\alpha c^{(\ell-1)})_k$. In particular, since $\mathcal{S}_\alpha$ is interpolating, we have that $d_{2k}^{(\ell)} = 0$ for all $k \in \mathbb{Z}$, that is,

$$\left[(\mathcal{I} - \mathcal{S}_\alpha \downarrow 2)c^{(\ell)}\right] \downarrow 2 = \mathbf{0}, \quad \ell = 1, 2, \ldots, J, \tag{10}$$

where $\mathcal{I}$ is the identity operator in the functional setting. Therefore, property (10) allows us to omit 'half' of the detail coefficients of each layer as we represent real-valued sequences—a natural benefit for data compression. The diagrams of Fig. 1 demonstrate the interpolating multiscale transforms like (8) and its inverse.

In fact, the interpolating multiscale transform (8) is a special case of the family of transforms presented in Donoho (1992). In particular, the operators $\mathcal{S}_{\boldsymbol{\alpha}}$ and $\downarrow 2$ play the roles of *upscaling* and *downscaling* filters, respectively.

2.4   *Noninterpolating linear multiscale transform*

The difficulty in using noninterpolating upscaling operators $\mathcal{S}_{\boldsymbol{\alpha}}$ in multiscale like (8) is that the sequence $\mathcal{S}_{\boldsymbol{\alpha}}(\boldsymbol{c})$ does not preserve the elements $\boldsymbol{c}$. In such case, the details must include more than just the difference between the original sequence $\boldsymbol{c}$ and refined downsampled sequence $\mathcal{S}_{\boldsymbol{\alpha}}(\boldsymbol{c} \downarrow 2)$.

The extension of multiscale transforms from interpolating subdivision operators to a wider class of subdivision operators involves *even-reversible* operators. Each of these operators helps recover, after one iteration of refinement, data points associated with even indices. In other words, given a subdivision operator $\mathcal{S}_{\boldsymbol{\alpha}}$, we seek for an operator $\mathcal{D}$ such that

$$\left[(\mathcal{I} - \mathcal{S}_{\boldsymbol{\alpha}} \mathcal{D}) \boldsymbol{c} \right] \downarrow 2 = \boldsymbol{0} \tag{11}$$

holds for any real-valued sequence $\boldsymbol{c}$. Indeed, condition (11) is the analogue of condition (10). Specifically, $\downarrow 2$ is replaced with the operator $\mathcal{D}$.

Let $\boldsymbol{\gamma}$ be a sequence such that $\sum_{i \in \mathbb{Z}} \gamma_i = 1$. Then, for any real-valued sequence $\boldsymbol{c}$ we define the *decimation operator* $\mathcal{D}_{\boldsymbol{\gamma}}$ associated with the sequence $\boldsymbol{\gamma}$ to be

$$\mathcal{D}_{\boldsymbol{\gamma}}(\boldsymbol{c})_k = \sum_{i \in \mathbb{Z}} \gamma_{k-i} c_{2i}, \quad k \in \mathbb{Z}. \tag{12}$$

Indeed, decimation operators are downscaling operators in the sense that applying them to a sequence of data results in fewer data. Note that $\boldsymbol{\gamma}$ can have infinite support. Thus, calculating (12) usually involves truncation errors. The rule (12) can be expressed as the unique solution of an optimization problem as similar to (5), where α_{k-2i} and c_i are replaced with γ_{k-i} and c_{2i}, respectively. Moreover, it can also be expressed in terms of the convolutional equation,

$$\mathcal{D}_{\boldsymbol{\gamma}}(\boldsymbol{c}) = \boldsymbol{\gamma} * (\boldsymbol{c} \downarrow 2).$$

Note that in the interpolation case, $\mathcal{D}_{\boldsymbol{\delta}}$ agrees with $\downarrow 2$ where $\boldsymbol{\delta}$ is the Kronecker delta sequence, $\delta_0 = 1$ and $\delta_i = 0$ for $i \neq 0$.

The unique solution of (11), see Dyn & Zhuang (2020), is the decimation operator $\mathcal{D}_{\boldsymbol{\gamma}}$ where $\boldsymbol{\gamma}$ is found via the convolutional equation

$$\boldsymbol{\gamma} * (\boldsymbol{\alpha} \downarrow 2) = \boldsymbol{\delta}. \tag{13}$$

Using Wiener's Lemma (Gröchenig, 2010), if $\boldsymbol{\alpha} \downarrow 2$ is compactly supported, then such $\boldsymbol{\gamma}$ with infinite support exists. In this case we say that $\mathcal{D}_{\boldsymbol{\gamma}}$ is the even-inverse of $\mathcal{S}_{\boldsymbol{\alpha}}$. Furthermore, $\boldsymbol{\gamma}$ decays geometrically, as shown in Strohmer (2002). More precisely,

$$|\gamma_k| \leq C \lambda^{|k|}, \quad k \in \mathbb{Z}, \tag{14}$$

for constants $C > 0$ and $0 < \lambda < 1$. This bound on the decay rate is essential for the computation of the decimation operation $\mathcal{D}_{\gamma}$, as we will see in the next section. We proceed with two examples of subdivision schemes that generate B-spline curves. First, we invoke the general formula of their compact masks. The mask $\boldsymbol{\alpha}^{[m]}$ of the B-spline subdivision operator of order $m \in \mathbb{N}$ is given by

$$\boldsymbol{\alpha}^{[m]}_{k-\lceil m/2 \rceil} = 2^{-m} \binom{m+1}{k}, \quad k = 0, 1, \ldots, m+1.$$

For more details see Dyn (1992).

EXAMPLE 2.1 *(The quadratic B-spline).* Consider the mask $\boldsymbol{\alpha}^{[2]} = [\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2] = \frac{1}{4}[1, 3, 3, 1]$, then the downsampled mask is $\boldsymbol{\alpha} \downarrow 2 = \frac{1}{4}[3, 1]$ and the solution of the corresponding convolutional equation (13) is

$$\gamma_k = \begin{cases} \dfrac{4}{3}\left(-\dfrac{1}{3}\right)^k, & k = 0, 1, 2, \ldots, \\ 0, & \text{otherwise.} \end{cases}$$

This subdivision scheme is also known as the *corner-cutting* scheme.

EXAMPLE 2.2 *(The cubic B-spline).* The next scheme generates cubic B-splines and its mask is given as $\boldsymbol{\alpha}^{[3]} = [\alpha_{-2}, \alpha_{-1}, \alpha_0, \alpha_1, \alpha_2] = \frac{1}{8}[1, 4, 6, 4, 1]$, then the downsampled mask is $\boldsymbol{\alpha} \downarrow 2 = \frac{1}{8}[1, 6, 1]$ and the solution of the corresponding convolutional equation (13) is

$$\begin{aligned}
&\left[\ldots, \gamma_{-3}, \gamma_{-2}, \gamma_{-1}, \gamma_0, \gamma_1, \gamma_2, \gamma_3 \ldots\right] \\
&= \left[\ldots, -0.0071, 0.0416, -0.2426, 1.4142, -0.2426, 0.0416, -0.0071 \ldots\right].
\end{aligned}$$

Note that in Examples 2.1 and 2.2 $\boldsymbol{\gamma}$, is infinitely and bi-infinitely supported, respectively.

We are finally in a position to present the noninterpolating linear multiscale transform. Given a noninterpolating subdivision scheme $\mathcal{S}_{\boldsymbol{\alpha}}$ with its corresponding even-inverse decimation operator $\mathcal{D}_{\boldsymbol{\gamma}}$, and a sequence $\boldsymbol{c}^{(J)}$ associated with the values over the fine grid $2^{-J}\mathbb{Z}$, we consider the following noninterpolating multiscale transform,

$$\boldsymbol{c}^{(\ell-1)} = \mathcal{D}_{\boldsymbol{\gamma}}\boldsymbol{c}^{(\ell)}, \quad \boldsymbol{d}^{(\ell)} = \boldsymbol{c}^{(\ell)} - \mathcal{S}_{\boldsymbol{\alpha}}\boldsymbol{c}^{(\ell-1)}, \quad \ell = 1, 2, \ldots, J. \tag{15}$$

Iterating (15) yields a pyramid of data $\left\{\boldsymbol{c}^{(0)}; \boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(J)}\right\}$ as similar to the interpolating transform (8). We obtain synthesis again by (9).

It turns out that the multiscale transform (15) enjoys two main properties, which are decay of the detail coefficients and stability of the inverse transform. Here we invoke both results citing Dyn & Zhuang (2020), but first, we define the operator Δ on a sequence $\boldsymbol{c}$ to be $\Delta \boldsymbol{c} = \sup_{k \in \mathbb{Z}} |c_{k+1} - c_k|$, and call decimation sequences $\boldsymbol{\gamma}$ that sum to 1 as shift invariant. Furthermore, we recall that the ∞-norm of a real-valued sequence $\boldsymbol{v} = \{v_k \mid k \in \mathbb{Z}\}$ is defined by $\|\boldsymbol{v}\|_{\infty} = \sup_{k \in \mathbb{Z}} |v_k|$.

THEOREM 2.3 Let $\boldsymbol{c}^{(J)}$ be a real-valued sequence and denote by $\left\{\boldsymbol{c}^{(0)}; \boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(J)}\right\}$ its multiscale transform generated by (15). The linear subdivision scheme is $\mathcal{S}_{\boldsymbol{\alpha}}$ and $\mathcal{D}_{\boldsymbol{\gamma}}$ is its corresponding

decimation operator defined using a shift invariant sequence $\boldsymbol{\gamma}$. We further assume that $K_{\boldsymbol{\gamma}} = 2 \sum_{i \in \mathbb{Z}} |\gamma_i| |i| < \infty$. Then,

$$\|\boldsymbol{d}^{(\ell)}\|_{\infty} \leq K_{\boldsymbol{\alpha}, \boldsymbol{\gamma}} \Delta \boldsymbol{c}^{(\ell)}, \quad \ell = 1, 2, \ldots, J, \tag{16}$$

with $K_{\boldsymbol{\alpha}, \boldsymbol{\gamma}} = K_{\boldsymbol{\gamma}} \|\boldsymbol{\alpha}\|_1 + K_{\boldsymbol{\alpha}} \|\boldsymbol{\gamma}\|_1$ where $K_{\boldsymbol{\alpha}} = \sum_{i \in \mathbb{Z}} |\alpha_i| |i|$.

THEOREM 2.4   Let $\{\boldsymbol{c}^{(0)}; \boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(J)}\}$ and $\{\widetilde{\boldsymbol{c}}^{(0)}; \widetilde{\boldsymbol{d}}^{(1)}, \ldots, \widetilde{\boldsymbol{d}}^{(J)}\}$ be two pyramids of sequences. Then, there exists $L \geq 0$ such that

$$\|\boldsymbol{c}^{(J)} - \widetilde{\boldsymbol{c}}^{(J)}\|_{\infty} \leq L \left(\|\boldsymbol{c}^{(0)} - \widetilde{\boldsymbol{c}}^{(0)}\|_{\infty} + \sum_{i=1}^{J} \|\boldsymbol{d}^{(i)} - \widetilde{\boldsymbol{d}}^{(i)}\|_{\infty} \right),$$

where $\boldsymbol{c}^{(J)}$ and $\widetilde{\boldsymbol{c}}^{(J)}$ are reconstructed from their respective data pyramids via (9).

3. Approximated linear decimation

The multiscale transform (15) involves applying $\boldsymbol{\gamma}$, which is infinite. In this section we develop transforms with finitely supported coefficients, which are essential in practice. We derive the decay rates of the new multiscale schemes, and compare them with the decay rate of the original multiscale transform.

3.1   *Truncation of the decimation coefficients*

We approximate the operation $\mathcal{D}_{\boldsymbol{\gamma}}$ as defined in (12) by a proper truncation of $\boldsymbol{\gamma}$. Given a *truncation parameter* $\varepsilon > 0$, we define

$$\widetilde{\gamma}_k(\varepsilon) = \begin{cases} \gamma_k, & |\gamma_k| > \varepsilon, \\ 0, & \text{elsewhere.} \end{cases} \tag{17}$$

The bound (14) implies that the support of $\widetilde{\boldsymbol{\gamma}}(\varepsilon)$, which we denote by Ω_ε, is finite for any ε. For simplicity, since we assume ε is fixed, we omit the parameter from the sequence $\widetilde{\boldsymbol{\gamma}}$.

The next theorem provides an upper bound for detail coefficients that are generated by (15), but with $\mathcal{D}_{\widetilde{\boldsymbol{\gamma}}}$ as its decimation operator. Namely, given a real-valued sequence $\boldsymbol{c}^{(J)}, J \in \mathbb{N}$, we consider

$$\boldsymbol{c}^{(\ell-1)} = \mathcal{D}_{\widetilde{\boldsymbol{\gamma}}} \boldsymbol{c}^{(\ell)}, \quad \boldsymbol{d}^{(\ell)} = \boldsymbol{c}^{(\ell)} - \mathcal{S}_{\boldsymbol{\alpha}} \boldsymbol{c}^{(\ell-1)}, \quad \ell = 1, 2, \ldots, J. \tag{18}$$

THEOREM 3.1   Let $\boldsymbol{c}^{(J)}$ be a real-valued sequence and denote by $\{\boldsymbol{c}^{(0)}; \boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(J)}\}$ its multiscale transform generated by (18). The subdivision scheme $\mathcal{S}_{\boldsymbol{\alpha}}$ is noninterpolating and $\mathcal{D}_{\widetilde{\boldsymbol{\gamma}}}$ is its corresponding decimation operator (12) with the truncated mask $\widetilde{\boldsymbol{\gamma}}$ of (17) where $\boldsymbol{\gamma}$ solves (13). Then,

$$\|\boldsymbol{d}^{(\ell)}\|_{\infty} \leq K_{\boldsymbol{\alpha}, \widetilde{\boldsymbol{\gamma}}} \Delta \boldsymbol{c}^{(\ell)} + \eta \|\boldsymbol{\alpha}\|_1 \|\boldsymbol{c}^{(\ell)}\|_{\infty}, \quad \ell = 1, 2, \ldots, J, \tag{19}$$

where $K_{\boldsymbol{\alpha}, \widetilde{\boldsymbol{\gamma}}} = K_{\widetilde{\boldsymbol{\gamma}}} \|\boldsymbol{\alpha}\|_1 + M K_{\boldsymbol{\alpha}}$ with $\eta = \sum_{i \notin \Omega_\varepsilon} |\gamma_i|, M = \sum_i |\widetilde{\gamma}_i|$ and $K_{\widetilde{\boldsymbol{\gamma}}} = 2 \sum_i |\widetilde{\gamma}_i| |i|$.

Proof. First, we calculate a general term in $\boldsymbol{d}^{(\ell)}$. For $k \in \mathbb{Z}$ we have

$$
\begin{aligned}
d_k^{(\ell)} &= c_k^{(\ell)} - \sum_i \alpha_{k-2i}(c^{(\ell-1)})_i = \sum_i \alpha_{k-2i}(c_k^{(\ell)} - c_i^{(\ell-1)}) = \sum_i \alpha_{k-2i}\left(c_k^{(\ell)} - \sum_n \widetilde{\gamma}_{i-n} c_{2n}^{(\ell)}\right) \\
&= \sum_i \alpha_{k-2i}\left(\sum_n \gamma_{i-n} c_k^{(\ell)} - \sum_{\substack{n \\ i-n \in \Omega_\varepsilon}} \gamma_{i-n} c_{2n}^{(\ell)}\right) \\
&= \sum_i \alpha_{k-2i}\left(\sum_{\substack{n \\ i-n \in \Omega_\varepsilon}} \gamma_{i-n}(c_k^{(\ell)} - c_{2n}^{(\ell)}) + \sum_{\substack{n \\ i-n \notin \Omega_\varepsilon}} \gamma_{i-n} c_k^{(\ell)}\right).
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
\|\boldsymbol{d}^{(\ell)}\|_\infty &\leq \sum_i |\alpha_{k-2i}|\left(\sum_{\substack{n \\ i-n \in \Omega_\varepsilon}} |\gamma_{i-n}| \cdot |c_k^{(\ell)} - c_{2n}^{(\ell)}| + \sum_{\substack{n \\ i-n \notin \Omega_\varepsilon}} |\gamma_{i-n}| \cdot |c_k^{(\ell)}|\right) \\
&\leq \sum_i |\alpha_{k-2i}|\left(\sum_{\substack{n \\ i-n \in \Omega_\varepsilon}} |\gamma_{i-n}| \cdot |2n - k| \cdot \Delta c^{(\ell)} + \eta |c_k^{(\ell)}|\right) \\
&\leq \sum_i |\alpha_{k-2i}|\left(\sum_{\substack{n \\ i-n \in \Omega_\varepsilon}} |\gamma_{i-n}| \cdot (|2n - 2i| + |k - 2i|) \cdot \Delta c^{(\ell)} + \eta \|c^{(\ell)}\|_\infty\right) \\
&\leq \sum_i |\alpha_{k-2i}|\left((K_{\widetilde{\boldsymbol{\gamma}}} + M|k - 2i|) \cdot \Delta c^{(\ell)} + \eta \|c^{(\ell)}\|_\infty\right) \\
&\leq (K_{\widetilde{\boldsymbol{\gamma}}}\|\boldsymbol{\alpha}\|_1 + M K_{\boldsymbol{\alpha}})\Delta c^{(\ell)} + \eta \|\boldsymbol{\alpha}\|_1 \|c^{(\ell)}\|_\infty.
\end{aligned}
$$

□

The term $\eta \|\boldsymbol{\alpha}\|_1 \|c^{(\ell)}\|_\infty$ in Theorem 3.1 is a direct result of the truncation (17). In particular, comparing with Theorem 2.3, $\varepsilon \to 0^+$ implies that $\Omega_\varepsilon \to \mathbb{Z}, \widetilde{\boldsymbol{\gamma}} \to \boldsymbol{\gamma}, \eta \to 0, K_{\widetilde{\boldsymbol{\gamma}}} \to K_{\boldsymbol{\gamma}}, M \to \|\boldsymbol{\gamma}\|_1$, and thus $K_{\boldsymbol{\alpha},\widetilde{\boldsymbol{\gamma}}} \to K_{\boldsymbol{\alpha},\boldsymbol{\gamma}}$. Consequently, by substituting the terms in (19), we obtain the exact bound as it appears in Theorem 2.3. More on the term $\eta \|\boldsymbol{\alpha}\|_1 \|c^{(\ell)}\|_\infty$, see Section 6.1.

3.2 *Normalization of the truncated coefficients*

Motivated by the case of manifold-valued data and following constructions of manifold-valued subdivision schemes, we require the truncated mask $\widetilde{\boldsymbol{\gamma}}$ to be shift invariant.

DEFINITION 3.2 Given a finitely supported mask $\widetilde{\boldsymbol{\gamma}}$ as in (17), we define $\boldsymbol{\zeta}$ to be the following normalized mask,

$$
\zeta_k = \frac{\widetilde{\gamma}_k}{\sum_{i \in \Omega_\varepsilon} \widetilde{\gamma}_i}, \quad k \in \mathbb{Z}. \tag{20}
$$

As it turns out the normalized truncated sequence $\boldsymbol{\zeta}$ of Definition 3.2 directly affects the decay rate of the detail coefficients. Let $\mathcal{S}_{\boldsymbol{\alpha}}$ be a subdivision scheme, and let $\mathcal{D}_{\boldsymbol{\zeta}}$ be its corresponding even-inverse decimation operator associated with the normalized truncated mask $\boldsymbol{\zeta}$ as defined in (20). Then, we define the multiscale transform,

$$\boldsymbol{c}^{(\ell-1)} = \mathcal{D}_{\boldsymbol{\zeta}} \boldsymbol{c}^{(\ell)}, \quad \boldsymbol{d}^{(\ell)} = \boldsymbol{c}^{(\ell)} - \mathcal{S}_{\boldsymbol{\alpha}} \boldsymbol{c}^{(\ell-1)}, \quad \ell = 1, 2, \ldots, J. \tag{21}$$

The following theorem shows that the sup norms of the detail coefficients generated by (21) are proportional to $\Delta \boldsymbol{c}^{(\ell)}$.

THEOREM 3.3 Let $\boldsymbol{c}^{(J)}$ be a real-valued sequence and denote by $\{\boldsymbol{c}^{(0)}; \boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(J)}\}$ its multiscale transform generated by (21). The subdivision scheme is $\mathcal{S}_{\boldsymbol{\alpha}}$ and $\mathcal{D}_{\boldsymbol{\zeta}}$ is its corresponding decimation operator with the normalized mask $\boldsymbol{\zeta}$ of (20) where $\boldsymbol{\gamma}$ solves (13). Then,

$$\|\boldsymbol{d}^{(\ell)}\|_{\infty} \leq K_{\boldsymbol{\alpha}, \boldsymbol{\zeta}} \Delta \boldsymbol{c}^{(\ell)}, \quad \ell = 1, 2, \ldots, J, \tag{22}$$

where $K_{\boldsymbol{\alpha}, \boldsymbol{\zeta}} = K_{\boldsymbol{\zeta}} \|\boldsymbol{\alpha}\|_1 + M K_{\boldsymbol{\alpha}}$ with $M = \sum_{i \in \Omega_\varepsilon} |\zeta_i|$ and $K_{\boldsymbol{\zeta}} = 2 \sum_{i \in \Omega_\varepsilon} |\zeta_i||i|$.

Proof. First, we calculate a general term in $\boldsymbol{d}^{(\ell)}$. For $k \in \mathbb{Z}$ we have

$$d_k^{(\ell)} = c_k^{(\ell)} - \sum_i \alpha_{k-2i} c_i^{(\ell-1)} = \sum_i \alpha_{k-2i} (c_k^{(\ell)} - c_i^{(\ell-1)}) = \sum_i \alpha_{k-2i} \left(c_k^{(\ell)} - \sum_n \zeta_{i-n} c_{2n}^{(\ell)} \right)$$

$$= \sum_i \alpha_{k-2i} \left(\sum_n \zeta_{i-n} c_k^{(\ell)} - \sum_n \zeta_{i-n} c_{2n}^{(\ell)} \right) = \sum_i \alpha_{k-2i} \left(\sum_n \zeta_{i-n} (c_k^{(\ell)} - c_{2n}^{(\ell)}) \right).$$

Consequently, similar arguments used in the proof of Theorem 3.1 yield to

$$\|\boldsymbol{d}^{(\ell)}\|_{\infty} \leq \sum_i |\alpha_{k-2i}| \left(K_{\boldsymbol{\zeta}} + M \cdot |k - 2i| \right) \Delta \boldsymbol{c}^{(\ell)}$$

$$\leq \left(K_{\boldsymbol{\zeta}} \|\boldsymbol{\alpha}\|_1 + M K_{\boldsymbol{\alpha}} \right) \cdot \Delta \boldsymbol{c}^{(\ell)} = K_{\boldsymbol{\alpha}, \boldsymbol{\zeta}} \Delta \boldsymbol{c}^{(\ell)},$$

as required.   □

To realize the importance of the normalization, we estimate the magnitude of the term $\Delta \boldsymbol{c}^{(\ell)}$ in (22) with respect to the level ℓ. This is achieved by assuming a prior on the sequence $\boldsymbol{c}^{(J)}$, as the following lemma suggests.

LEMMA 3.4 Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable, bounded real-valued function. Denote by $\boldsymbol{c}^{(J)}, J \in \mathbb{N}$ the function's samples over the grid $2^{-J} \mathbb{Z}$, that is, $\boldsymbol{c}^{(J)} = f |_{2^{-J} \mathbb{Z}}$, and let $\mathcal{D}_{\boldsymbol{\zeta}}$ be a decimation operator (12) associated with the shift invariant sequence $\boldsymbol{\zeta}$. Then,

$$\Delta \boldsymbol{c}^{(\ell)} \leq \|\boldsymbol{\zeta}\|_1^J \cdot \|f'\|_{\infty} \cdot (2\|\boldsymbol{\zeta}\|_1)^{-\ell}, \quad \ell = 0, 1, \ldots, J, \tag{23}$$

where $\|f'\|_{\infty} = \sup_{x \in \mathbb{R}} |f'(x)|$ and the sequences $\boldsymbol{c}^{(\ell)}$ are generated iteratively by (21).

Proof. Since f is differentiable and bounded, then by the mean value theorem, for all $k \in \mathbb{Z}$ and a fixed $J \in \mathbb{N}$, there exists x_k in the open segment, which connects the parametrizations of $c_k^{(J)}$ and $c_{k+1}^{(J)}$, such that

$$|c_{k+1}^{(J)} - c_k^{(J)}| = 2^{-J}|f'(x_k)|,$$

and by applying the supremum over all $k \in \mathbb{Z}$ we obtain $\Delta c^{(J)} \leq 2^{-J}\|f'\|_\infty$. Now, observe that for any real-valued sequence $\boldsymbol{c}$, we have

$$\Delta(\boldsymbol{c} \downarrow 2) \leq \sup_{k \in \mathbb{Z}} |c_{2k+2} - c_{2k}| \leq 2 \cdot \sup_{k \in \mathbb{Z}} |c_{k+1} - c_k| = 2\Delta \boldsymbol{c},$$

and since the convolution commutes with Δ, we get

$$
\begin{aligned}
\Delta \boldsymbol{c}^{(\ell-1)} &= \Delta\big(\boldsymbol{\zeta} * (\boldsymbol{c}^{(\ell)} \downarrow 2)\big) \\
&\leq \|\boldsymbol{\zeta}\|_1 \cdot \Delta(\boldsymbol{c}^{(\ell)} \downarrow 2) \leq 2\|\boldsymbol{\zeta}\|_1 \cdot \Delta \boldsymbol{c}^{(\ell)}.
\end{aligned}
\tag{24}
$$

Iterating the latter inequality starting with ℓ gives

$$\Delta \boldsymbol{c}^{(\ell)} \leq 2\|\boldsymbol{\zeta}\|_1 \cdot \Delta \boldsymbol{c}^{(\ell+1)} \leq (2\|\boldsymbol{\zeta}\|_1)^2 \cdot \Delta \boldsymbol{c}^{(\ell+2)} \leq \cdots \leq (2\|\boldsymbol{\zeta}\|_1)^{J-\ell} \cdot \Delta \boldsymbol{c}^{(J)},$$

which is equivalent to (23), for any $\ell = 0, 1, \ldots, J$. $\qquad\square$

Theorem 3.3 and Lemma 3.4 illustrate the significance of the normalization (20). Specifically, if $\boldsymbol{c}^{(J)}$ is sampled from a differentiable function, then the detail coefficients generated by the multiscale transform (21) are bounded by a geometrically decreasing bound, as the following corollary states.

COROLLARY 3.5 If $\boldsymbol{c}^{(J)}$ are sampled from a differentiable function $f \colon \mathbb{R} \to \mathbb{R}$ over the grid $2^{-J}\mathbb{Z}$, then the detail coefficients generated by (21) satisfy

$$\|\boldsymbol{d}^{(\ell)}\|_\infty \leq K_{\boldsymbol{\alpha},\boldsymbol{\zeta}} \|\boldsymbol{\zeta}\|_1^J \|f'\|_\infty \cdot (2\|\boldsymbol{\zeta}\|_1)^{-\ell}, \quad \ell = 1, 2, \ldots, J. \tag{25}$$

The upper bound in (25) decays geometrically with factor $1/(2\|\boldsymbol{\zeta}\|_1) < 1/2$, with respect to the level ℓ.

A direct result from Corollary 3.5 suggests that the detail coefficients corresponding to the highest scale can be as small as we desire. In particular, for any $\sigma > 0$, one can find a sufficiently large $J_0 \in \mathbb{N}$ such that $\|\boldsymbol{d}^{(J)}\|_\infty \leq \sigma$ for all $J > J_0$. A direct calculation for the minimal J_0 yields

$$J_0 = \big\lfloor \log_2 \big(\|f'\|_\infty \cdot K_{\boldsymbol{\alpha},\boldsymbol{\zeta}}/\sigma\big) \big\rfloor + 1, \tag{26}$$

in contrary to the transform (18) where such J_0 is not guaranteed. Note that according to (26), when $\|f'\|_\infty$ is large, for example, if f is rapidly changing, we need a denser grid to achieve small enough details, that is $\|\boldsymbol{d}^{(J)}\|_\infty \leq \sigma$.

4. Noninterpolating transform for manifold-valued data

Interpolating multiscale transforms have been studied in various nonlinear settings (Grohs & Wallner, 2009, 2012; Kaur & Singh, 2017; Lv *et al.*, 2018). In this section, we aim to adapt the noninterpolating multiscale transform (21) to manifold-valued data.

4.1 *The Riemannian analogue of a linear decimation operator*

Let $\boldsymbol{\zeta}$ be a normalized finitely supported mask (20), and let $\mathcal{M}$ be a Riemannian manifold. We extend the decimation operator (12) to $\mathcal{M}$ with the help of its Riemannian geodesic distance (4), similar to the extension of the subdivision schemes, as done in Section 2.2.

For any $\mathcal{M}$-valued sequence $\boldsymbol{c}$, we define

$$\mathcal{Y}_{\boldsymbol{\zeta}}(\boldsymbol{c})_k = \arg\min_{x \in \mathcal{M}} \sum_{i \in \mathbb{Z}} \zeta_{k-i} \rho(x, c_{2i})^2, \quad k \in \mathbb{Z}. \tag{27}$$

Namely, $\mathcal{Y}_{\boldsymbol{\zeta}}(\boldsymbol{c})_k$ is interpreted as the Riemannian center of mass of the elements c_{2i} with the corresponding weights ζ_{k-i}. Moreover, it is the Riemannian analogue of the linear decimation operator $\mathcal{D}_{\boldsymbol{\zeta}}$. Results in Hardering (2015) provide conditions for the global existence of a minimizer in (27). We formulate one such result in the following lemma.

LEMMA 4.1 Let $\boldsymbol{\zeta} \in \mathbb{R}^m$ be a shift invariant mask, and let $\boldsymbol{c} = \{c_1, c_2, \ldots, c_m\}$ be a set of $\mathcal{M}$-valued points, satisfying $\rho(c_1, c_k) \leq r$ for some $r > 0$ and for all $k = 1, 2, \ldots, m$. Then, the objective function $h \colon \mathcal{M} \to \mathbb{R}$ defined as,

$$h(x) = \sum_{i=1}^{m} \zeta_i \rho(x, c_i)^2 \tag{28}$$

has at least one minimum. Moreover, there exists a constant $R \leq 6m\|\boldsymbol{\zeta}\|_{\infty}$ such that all minima of (28) lie inside a compact ball centred around c_1 with radius rR. If r is small enough, with respect to the curvature of $\mathcal{M}$, then (28) has a unique solution.

Henceforth, we require any $\mathcal{M}$-valued admissible sequences to obey the strong conditions of Lemma (4.1) for any shift invariant mask of our decimation operator $\mathcal{Y}_{\boldsymbol{\zeta}}$, that is, the Riemannian center of mass exists and unique. Moreover, we say that the decimation operator $\mathcal{Y}_{\boldsymbol{\zeta}}$ of (27) is the even-inverse of the subdivision scheme $\mathcal{T}_{\boldsymbol{\alpha}}$ of (6) if its linear version $\mathcal{D}_{\boldsymbol{\zeta}}$, associated with the normalized mask $\boldsymbol{\zeta}$ of (20), where $\boldsymbol{\gamma}$ solves (13), is the even-inverse of the linear scheme $\mathcal{S}_{\boldsymbol{\alpha}}$. We illustrate the process of adaptation of the decimation operator to manifold values in the diagram of Fig. 2.

4.2 *The noninterpolating multiscale transform for manifold data*

Now that we have the adapted subdivision operator $\mathcal{T}_{\boldsymbol{\alpha}}$ and its corresponding decimation operator $\mathcal{Y}_{\boldsymbol{\zeta}}$ being defined, we proceed by adjusting the linear pyramid transform (21) to manifold values. A first difference between the manifold and linear versions of the transform lies in the coefficients. For a manifold-valued transform the sequence $\boldsymbol{c}^{(\ell)}$ at level ℓ is a $\mathcal{M}$-valued sequence, while the detail coefficients $\boldsymbol{d}^{(\ell)}$ are elements in the tangent bundle $T\mathcal{M} = \bigcup_{p \in \mathcal{M}} \{p\} \times T_p\mathcal{M}$ associated with $\mathcal{M}$.

Recall that in a Riemannian manifold $\mathcal{M}$, the exponential mapping $\exp_p$ maps a vector v in the tangent space $T_p\mathcal{M}$ to the end point of a geodesic of length $\|v\|$, which emanates from $p \in \mathcal{M}$ with
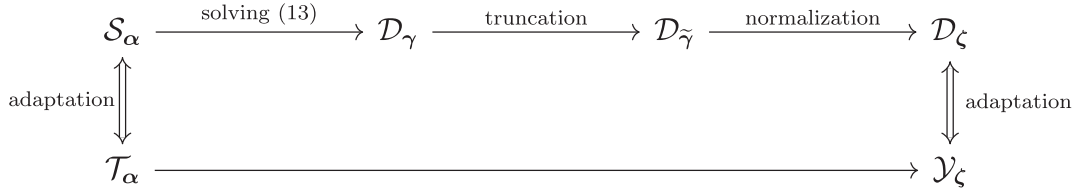
$$\mathcal{S}_{\boldsymbol{\alpha}} \xrightarrow{\text{solving (13)}} \mathcal{D}_{\gamma} \xrightarrow{\text{truncation}} \mathcal{D}_{\widetilde{\gamma}} \xrightarrow{\text{normalization}} \mathcal{D}_{\zeta}$$

adaptation $\Updownarrow$ $\Updownarrow$ adaptation

$$\mathcal{T}_{\boldsymbol{\alpha}} \xrightarrow{\hspace{8cm}} \mathcal{Y}_{\zeta}$$

Fig. 2. The upper and lower rows represent linear and nonlinear operators, respectively. The process of finding the even-inverse of the manifold-valued subdivision scheme $\mathcal{T}_{\boldsymbol{\alpha}}$ is done indirectly by going through the linear operators and adapting the resulting decimation operator.

initial tangent vector v. Inversely, $\log_p$ is the inverse map of $\exp_p$ that takes an $\mathcal{M}$-valued element q and returns a vector in the tangent space $T_p\mathcal{M}$. Following similar notations used in Grohs & Wallner (2009), we denote both maps by

$$\log_p(q) = q \ominus p \quad \text{and} \quad \exp_p(v) = p \oplus v. \tag{29}$$

We have thus defined the analogues $\ominus$ and $\oplus$ of the ' $-$ ' and ' $+$ ' operations, respectively.

For any point $p \in \mathcal{M}$ we use the following notation $\ominus \colon \mathcal{M}^2 \to T_p\mathcal{M}$ and $\oplus \colon \mathcal{M} \times T_p\mathcal{M} \to \mathcal{M}$. Then, the compatibility condition is

$$(p \oplus v) \ominus p = v, \tag{30}$$

for all $v \in T_p\mathcal{M}$ within the injectivity radius of $\mathcal{M}$. That being, so we are now capable of introducing our noninterpolating multiscale transform for manifold-valued data. Let $\mathcal{M}$ be a Riemannian manifold, and let $\boldsymbol{c}^{(J)}$, $J \in \mathbb{N}$ be $\mathcal{M}$-valued sequence associated with the values over the grid $2^{-J}\mathbb{Z}$. Given a noninterpolating subdivision scheme $\mathcal{T}_{\boldsymbol{\alpha}}$ with its corresponding decimation operator $\mathcal{Y}_{\zeta}$, we define the noninterpolating multiscale transform

$$\boldsymbol{c}^{(\ell-1)} = \mathcal{Y}_{\zeta}\boldsymbol{c}^{(\ell)}, \quad \boldsymbol{d}^{(\ell)} = \boldsymbol{c}^{(\ell)} \ominus \mathcal{T}_{\boldsymbol{\alpha}}\boldsymbol{c}^{(\ell-1)}, \quad \ell = 1, 2, \ldots, J, \tag{31}$$

where the operation $\ominus$ is the log map (29) associated with $\mathcal{M}$.

Indeed, the transform (31) is the nonlinear analogue of (21). The process (31) yields a pyramid of sequences $\{\boldsymbol{c}^{(0)}; \boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(J)}\}$, where $\boldsymbol{c}^{(0)}$ are the coarse approximation coefficients given over the integers, and $\boldsymbol{d}^{(\ell)}$, $\ell = 1, 2, \ldots, J$ are the detail coefficients at level ℓ given over the values of the grids $2^{-\ell}\mathbb{Z}$, respectively. By the construction of (31), we verify that $\boldsymbol{c}^{(0)}$ is an $\mathcal{M}$-valued sequence and the elements of $\boldsymbol{d}^{(\ell)}$ lie in $T\mathcal{M}$ for all $\ell = 1, 2, \ldots, J$. Then, we obtain synthesis by the iterations,

$$\boldsymbol{c}^{(\ell)} = \mathcal{T}_{\boldsymbol{\alpha}}\boldsymbol{c}^{(\ell-1)} \oplus \boldsymbol{d}^{(\ell)}, \quad \ell = 1, 2, \ldots, J. \tag{32}$$

The synthesis (32) is the analogue of (9), and it is the inverse transform of (31).

5. Coefficients decay and reconstruction stability

A vital feature of any multiscale transform is the rate at which the detail coefficients become small and, therefore, from a certain point, negligible. This feature is also the basis of various thresholding

techniques for compression, smoothness analysis and denoising, as we will see in Section 6. This section shows that, under stability of subdivision schemes, the detail coefficients are bounded, and the inverse transform is stable.

5.1   *Displacement safe operators*

We proceed with a series of useful definitions. First, for admissible $\mathcal{M}$-valued sequences $\boldsymbol{c}$ and $\boldsymbol{m}$, we denote the following. Let

$$\Delta_{\mathcal{M}}(\boldsymbol{c}) = \sup_{k \in \mathbb{Z}} \rho(c_{k+1}, c_k) \quad \text{and} \quad \mu(\boldsymbol{c}, \boldsymbol{m}) = \sup_{k \in \mathbb{Z}} \rho(c_k, m_k),$$

be the supremum distance between consecutive elements in $\boldsymbol{c}$ and the corresponding elements of the sequences $\boldsymbol{c}$ and $\boldsymbol{m}$, respectively. We continue with the definition of displacement-safe subdivision schemes, as introduced in Dyn & Sharon (2017b). This condition means that the refinement operator generates points in a controlled fashion regarding the data points' distances. In linear schemes, for example, this condition is automatically satisfied.

DEFINITION 5.1   We say that the subdivision operator $\mathcal{T}$ is *displacement-safe* if there exists $E_{\mathcal{T}} \geq 0$ such that $\mu(\boldsymbol{c}, \mathcal{T}(\boldsymbol{c}) \downarrow 2) \leq E_{\mathcal{T}} \Delta_{\mathcal{M}}(\boldsymbol{c})$ for any $\mathcal{M}$-valued sequence $\boldsymbol{c}$.

Displacement safety plays a significant role for the convergence analysis of noninterpolating subdivision schemes over manifolds (Dyn & Sharon, 2017b; Hüning & Wallner, 2020). Next, we introduce a new condition analogous to Definition 5.1, but for decimation operators.

DEFINITION 5.2   We say that the decimation operator $\mathcal{Y}$ is *decimation-safe* if there exists $F_{\mathcal{Y}} \geq 0$ such that $\mu(\mathcal{Y}\boldsymbol{c}, \boldsymbol{c} \downarrow 2) \leq F_{\mathcal{Y}} \Delta_{\mathcal{M}}(\boldsymbol{c})$ for any $\mathcal{M}$-valued sequence $\boldsymbol{c}$.

To simplify the terminologies, we say that the pair $(\mathcal{T}_{\boldsymbol{\alpha}}, \mathcal{Y}_{\boldsymbol{\zeta}})$ is *safe* if simultaneously both the operator $\mathcal{T}_{\boldsymbol{\alpha}}$ is displacement-safe and $\mathcal{Y}_{\boldsymbol{\zeta}}$ is its corresponding decimation-safe decimation operator. With the definitions being stated, we note that the linear pyramid transforms (18) and (21) involve safe pairs of subdivision and decimation operators, as we state and prove in the following lemma.

LEMMA 5.3   In the Euclidean case the linear pair $(\mathcal{S}_{\boldsymbol{\alpha}}, \mathcal{D}_{\boldsymbol{\zeta}})$ is safe.

Proof.   The proof that $\mathcal{S}_{\boldsymbol{\alpha}}$ is displacement safe appears in Dyn & Sharon (2017b), we proceed with proving that linear decimation operators are decimation-safe. For any real-valued sequence $\boldsymbol{c}$ and index $k \in \mathbb{Z}$, we have

$$|(\mathcal{D}_{\boldsymbol{\zeta}}\boldsymbol{c})_k - (\boldsymbol{c} \downarrow 2)_k| = \big| \sum_{i \in \mathbb{Z}} \zeta_{k-i} c_{2i} - c_{2k} \big| = \big| \sum_{i \in \mathbb{Z}} \zeta_{k-i} c_{2i} - \sum_{i \in \mathbb{Z}} \zeta_{k-i} c_{2k} \big|$$

$$= \big| \sum_{i \in \mathbb{Z}} \zeta_{k-i} (c_{2i} - c_{2k}) \big| \leq 2 \sum_{i \in \mathbb{Z}} |\zeta_{k-i}| \cdot |k - i| \cdot \Delta\boldsymbol{c}.$$

Now, applying $\sup_{k \in \mathbb{Z}}$ on both sides gives

$$\|\mathcal{D}_{\boldsymbol{\zeta}}\boldsymbol{c} - \boldsymbol{c} \downarrow 2\|_{\infty} \leq F_{\mathcal{D}} \Delta\boldsymbol{c},$$

with $F_{\mathcal{D}} = \sup_{k \in \mathbb{Z}} 2 \sum_{i \in \mathbb{Z}} |\zeta_{k-i}| \cdot |k - i| < \infty$ since $\boldsymbol{\zeta}$ has a compact support.   □

Note that in the case where $\mathcal{T}_{\boldsymbol{\alpha}}$ is interpolating, the corresponding decimation operator $\mathcal{Y}_{\boldsymbol{\zeta}}$ is simply the downsampling operator $\downarrow 2$, see e.g., Grohs & Wallner (2009); Grohs (2010b). Thus, the pair $\left(\mathcal{T}_{\boldsymbol{\alpha}}, \mathcal{Y}_{\boldsymbol{\zeta}}\right)$ is safe. In particular, $E_{\mathcal{T}} = F_{\mathcal{Y}} = 0$. Otherwise, Lemma 4.1 guarantees that the pair $\left(\mathcal{T}_{\boldsymbol{\alpha}}, \mathcal{Y}_{\boldsymbol{\zeta}}\right)$ is safe, as the following proposition suggests.

PROPOSITION 5.4 Let $\mathcal{T}_{\boldsymbol{\alpha}}$ be a noninterpolating subdivision scheme over $\mathcal{M}$, the pair $\left(\mathcal{T}_{\boldsymbol{\alpha}}, \mathcal{Y}_{\boldsymbol{\zeta}}\right)$ is safe for any truncation parameter ε involved in determining the shift invariant mask $\boldsymbol{\zeta}$.

Proof. Here we only prove that $\mathcal{Y}_{\boldsymbol{\zeta}}$ is decimation safe, see Definition 5.2, the proof that $\mathcal{T}_{\boldsymbol{\alpha}}$ is displacement safe is similar. Let $\boldsymbol{c}$ be an admissible $\mathcal{M}$-valued sequence, satisfying

$$\mathcal{H} = \frac{\sup_{|i-j| \leq |\Omega_{\varepsilon}|} \rho(c_i, c_j)}{\Delta_{\mathcal{M}}(\boldsymbol{c})} < \infty,$$

where Ω_{ε} is the compact support of $\boldsymbol{\zeta}$, as in (17). We express $\boldsymbol{c}$ as a countable union of pairwise overlapping sets Ξ_k, $k \in \mathbb{Z}$, where Ξ_k consists of all the elements of $\boldsymbol{c}$ involved in calculating $\mathcal{Y}_{\boldsymbol{\zeta}}(\boldsymbol{c})_k$, see (27). Indeed, Ξ_k contains $|\Omega_{\varepsilon}|$-many elements including c_{2k}. Lemma 4.1 guarantees

$$\rho(\mathcal{Y}_{\boldsymbol{\zeta}}(\boldsymbol{c})_k, c_{2k}) \leq 6|\Omega_{\varepsilon}| \|\boldsymbol{\zeta}\|_{\infty} \cdot r_k,$$

where $r_k = \max_{c_j \in \Xi_k} \rho(c_{2k}, c_j)$. By the definition of $\mathcal{H}$ we have $r_k \leq \mathcal{H}\Delta_{\mathcal{M}}(\boldsymbol{c})$, $k \in \mathbb{Z}$. Thus, applying $\sup_{k \in \mathbb{Z}}$ yields $\mu(\mathcal{Y}_{\boldsymbol{\zeta}}\boldsymbol{c}, \boldsymbol{c} \downarrow 2) \leq F_{\mathcal{Y}}\Delta_{\mathcal{M}}(\boldsymbol{c})$, as required, with $F_{\mathcal{Y}} = 6|\Omega_{\varepsilon}| \|\boldsymbol{\zeta}\|_{\infty}\mathcal{H}$. $\qquad\square$

In the same manner of using Lemma 4.1 to prove Proposition 5.4, the former guarantees that for any admissible $\mathcal{M}$-valued sequence $\boldsymbol{c}$, we have

$$\Delta_{\mathcal{M}}(\mathcal{T}_{\boldsymbol{\alpha}}\boldsymbol{c}) \leq Q\Delta_{\mathcal{M}}(\boldsymbol{c}) \tag{33}$$

for some $Q > 0$. Estimate (33) is used in the proof of Proposition 5.6 in the next section. Later on, in the proof of Lemma 5.8, we will observe the analogue of (33), but for decimation operators.

5.2 *Decay of detail coefficients*

We proceed with defining a stable subdivision rule.

DEFINITION 5.5 We say that the subdivision operator $\mathcal{T}$ is stable if there exists $S_{\mathcal{T}} \geq 0$ such that $\mu(\mathcal{T}\boldsymbol{c}, \mathcal{T}\boldsymbol{m}) \leq S_{\mathcal{T}}\mu(\boldsymbol{c}, \boldsymbol{m})$ for all $\mathcal{M}$-valued sequences $\boldsymbol{c}$ and $\boldsymbol{m}$.

Stable subdivision schemes have been studied in Grohs (2010b). The next proposition helps in proving the following theorem, which is the analogue to Theorem 3.3.

PROPOSITION 5.6 For any admissible sequence $\boldsymbol{c}$, we have

$$\mu\left(\boldsymbol{c}, \mathcal{T}_{\boldsymbol{\alpha}}(\boldsymbol{c} \downarrow 2)\right) \leq (1 + 2E_{\mathcal{T}} + Q)\Delta_{\mathcal{M}}(\boldsymbol{c}),$$

where the constants $E_{\mathcal{T}}$ and Q are from Definition 5.1 and (33), respectively.

Proof. Recall that $\mu\big(c, \mathcal{T}_{\boldsymbol{\alpha}}(c \downarrow 2)\big) = \sup_{k \in \mathbb{Z}} \rho\big(c_k, \mathcal{T}_{\boldsymbol{\alpha}}(c \downarrow 2)_k\big)$. If $k = 2j$ is even, then by the displacement-safe inequality, see Definition 5.1, we have

$$\rho\big(c_{2j}, \mathcal{T}_{\boldsymbol{\alpha}}(c \downarrow 2)_{2j}\big) = \rho\big((c \downarrow 2)_j, \mathcal{T}_{\boldsymbol{\alpha}}(c \downarrow 2)_{2j}\big) \leq E_{\mathcal{T}}\Delta_{\mathcal{M}}(c \downarrow 2) \leq 2E_{\mathcal{T}}\Delta_{\mathcal{M}}(c).$$

Otherwise, we have

$$\rho\big(c_{2j+1}, \mathcal{T}_{\boldsymbol{\alpha}}(c \downarrow 2)_{2j+1}\big) \leq \rho(c_{2j+1}, c_{2j}) + \rho\big(c_{2j}, \mathcal{T}_{\boldsymbol{\alpha}}(c \downarrow 2)_{2j}\big) + \rho\big(\mathcal{T}_{\boldsymbol{\alpha}}(c \downarrow 2)_{2j}, \mathcal{T}_{\boldsymbol{\alpha}}(c \downarrow 2)_{2j+1}\big)$$

$$\leq (1 + 2E_{\mathcal{T}} + Q)\Delta_{\mathcal{M}}(c),$$

as required. □

THEOREM 5.7 Let $\mathcal{T}_{\boldsymbol{\alpha}}$ be a stable subdivision scheme, then there exists $K \geq 0$ such that the detail coefficients generated by (31) satisfy

$$\|d^{(\ell)}\|_{\infty} \leq K\Delta_{\mathcal{M}}(c^{(\ell)}), \quad \ell = 1, 2, \ldots, J, \tag{34}$$

where $\|d^{(\ell)}\|_{\infty} = \sup_{k \in \mathbb{Z}} \|d_k^{(\ell)}\|$.

Proof. First, observe that $\|d^{(\ell)}\|_{\infty} = \|c^{(\ell)} \ominus \mathcal{T}_{\boldsymbol{\alpha}}c^{(\ell-1)}\|_{\infty} = \mu(c^{(\ell)}, \mathcal{T}_{\boldsymbol{\alpha}}c^{(\ell-1)})$. Since the pair $\big(\mathcal{T}_{\boldsymbol{\alpha}}, \mathcal{Y}_{\boldsymbol{\zeta}}\big)$ is safe, see Proposition 5.4, then by the triangle inequality, we have

$$\mu(c^{(\ell)}, \mathcal{T}_{\boldsymbol{\alpha}}c^{(\ell-1)}) \leq \mu\big(c^{(\ell)}, \mathcal{T}_{\boldsymbol{\alpha}}(c^{(\ell)} \downarrow 2)\big) + \mu\big(\mathcal{T}_{\boldsymbol{\alpha}}(c^{(\ell)} \downarrow 2), \mathcal{T}_{\boldsymbol{\alpha}}\mathcal{Y}_{\boldsymbol{\zeta}}c^{(\ell)}\big)$$

$$\leq (1 + 2E_{\mathcal{T}} + Q)\Delta_{\mathcal{M}}(c^{(\ell)}) + S_{\mathcal{T}}\mu\big(c^{(\ell)} \downarrow 2, \mathcal{Y}_{\boldsymbol{\zeta}}c^{(\ell)}\big)$$

$$\leq (1 + 2E_{\mathcal{T}} + Q + S_{\mathcal{T}}F_{\mathcal{Y}})\Delta_{\mathcal{M}}(c^{(\ell)}),$$

as required where $K = 1 + 2E_{\mathcal{T}} + Q + S_{\mathcal{T}}F_{\mathcal{Y}}$. □

To proceed, we estimate the magnitude of $\Delta_{\mathcal{M}}(c^{(\ell)})$ in (34) by assuming a prior on the admissible data points $c^{(J)}$. Recall that if Γ is an $\mathcal{M}$-valued regular differentiable curve, then $\nabla\Gamma(x)$ denotes the intrinsic gradient of Γ at point $x \in \mathcal{M}$, i.e., the velocity vector of Γ at point $x \in \mathcal{M}$ lying in $T_x\mathcal{M}$ (do Carmo, 1992). The following lemma is the analogue to Lemma 3.4.

LEMMA 5.8 Let Γ be a regular differentiable curve over $\mathcal{M}$. Denote by $c^{(J)}, J \in \mathbb{N}$ the curve's samples over the arc-length parametrization grid $2^{-J}\mathbb{Z}$, that is, $c^{(J)} = \Gamma \mid_{2^{-J}\mathbb{Z}}$, and let $\mathcal{Y}_{\boldsymbol{\zeta}}$ be a decimation operator associated with the shift invariant mask $\boldsymbol{\zeta}$. Then, there exists $P > 1$, depending on $\boldsymbol{\zeta}$ and the curvature of $\mathcal{M}$, such that

$$\Delta_{\mathcal{M}}(c^{(\ell)}) \leq P^J\|\nabla\Gamma\|_{\infty} \cdot (2P)^{-\ell}, \quad \ell = 0, 1, \ldots, J, \tag{35}$$

where $\|\nabla\Gamma\|_{\infty} = \sup_s \|\nabla\Gamma(s)\|$ and the sequences $c^{(\ell)}$ are generated recursively by (31).

Proof. First, since $\mathcal{Y}_{\boldsymbol{\zeta}}$ is decimation safe, see Proposition 5.4, then for $\ell = 1, 2, \ldots, J$ and $k \in \mathbb{Z}$,

$$\rho\big((\mathcal{Y}_{\boldsymbol{\zeta}}c^{(\ell)})_{k+1}, (\mathcal{Y}_{\boldsymbol{\zeta}}c^{(\ell)})_k\big) \leq \rho\big((\mathcal{Y}_{\boldsymbol{\zeta}}c^{(\ell)})_{k+1}, c_{2k+2}\big) + \rho\big(c_{2k+2}, c_{2k}\big) + \rho\big(c_{2k}, (\mathcal{Y}_{\boldsymbol{\zeta}}c^{(\ell)})_k\big).$$

Applying $\sup_{k \in \mathbb{Z}}$ yields to

$$\Delta_{\mathcal{M}}(\mathcal{Y}_{\zeta} c^{(\ell)}) \leq 2F_{\mathcal{Y}} \Delta_{\mathcal{M}}(c^{(\ell)}) + 2\Delta_{\mathcal{M}}(c^{(\ell)}) = 2P \cdot \Delta_{\mathcal{M}}(c^{(\ell)}),$$

where $P = 1 + F_{\mathcal{Y}}$. In other words, $\Delta_{\mathcal{M}}(c^{(\ell-1)}) \leq 2P \cdot \Delta_{\mathcal{M}}(c^{(\ell)})$. Iteratively, we get

$$\Delta_{\mathcal{M}}(c^{(\ell)}) \leq (2P)^{J-\ell} \Delta_{\mathcal{M}}(c^{(J)}), \quad \ell = 0, 1, \ldots, J.$$

Now, since the sequence $c^{(J)}$ is sampled from a regular differentiable curve Γ over the arc-length parametrization grid $2^{-J}\mathbb{Z}$, we immediately have $\Delta_{\mathcal{M}}(c^{(J)}) \leq \|\nabla \Gamma\|_{\infty} 2^{-J}$. In total, inequality (35) follows. □

Inequality (35) is the analogue to the estimate in (23). The next corollary is the analogue of Corollary 3.5, it shows that if $c^{(J)}$ is sampled from a differentiable curve over the manifold, then the detail coefficients generated by (31) are bounded by a geometrically decreasing bound with factor $1/(2P) < 1/2$.

COROLLARY 5.9 Let $c^{(J)}$ denote the samples of a regular differentiable curve Γ over $\mathcal{M}$ on the equispaced arc-length parameterized grid $2^{-J}\mathbb{Z}$, where $\|\nabla \Gamma\|_{\infty}$ assumed to be finite. Let $\mathcal{T}_{\alpha}$ be stable subdivision scheme, then the detail coefficients generated by (31) satisfy

$$\|d^{(\ell)}\|_{\infty} \leq KP^J \|\nabla \Gamma\|_{\infty} \cdot (2P)^{-\ell}, \quad \ell = 1, 2, \ldots, J, \tag{36}$$

with the same values of K in (34) and P in (35). The upper bound in (36) decays geometrically with factor $1/(2P) < 1/2$, with respect to the level ℓ.

In inequality (36), if $\ell = J$ then $\|d^{(J)}\|_{\infty} \leq K\|\nabla \Gamma\|_{\infty} \cdot 2^{-J}$. Thus, Corollary 5.9 guarantees that for any $\sigma > 0$ there exists some $J_0 \in \mathbb{N}$ such that $\|d^{(J)}\|_{\infty} \leq \sigma$ for all $J > J_0$. The minimal J_0 is similar in its form to (26).

5.3 *Stability of inverse transform*

In this section we discuss the stability of the inverse multiscale transform (32). As a first conclusion we show that the stability of the subdivision operator $\mathcal{T}_{\alpha}$, as presented in Definition 5.5, induces a stability result for the inverse transform.

THEOREM 5.10 Let $\{c^{(0)}; d^{(1)}, \ldots, d^{(J)}\}$ and $\{\widetilde{c}^{(0)}; \widetilde{d}^{(1)}, \ldots, \widetilde{d}^{(J)}\}$ be two pyramids of sequences. Let $\mathcal{T}_{\alpha}$ be stable with constant $S_{\mathcal{T}}$, as in Definition 5.5. Then, the synthesis sequences $c^{(J)}$ and $\widetilde{c}^{(J)}$, which are reconstructed from the above two data pyramids via (32), satisfy

$$\mu(c^{(J)}, \widetilde{c}^{(J)}) \leq L\left(\mu(c^{(0)}, \widetilde{c}^{(0)}) + \sum_{i=1}^{J} \|d^{(i)}\|_{\infty} + \|\widetilde{d}^{(i)}\|_{\infty}\right), \tag{37}$$

with $L = 1$ if $S_{\mathcal{T}} \leq 1$ and $L = S_{\mathcal{T}}^J$ otherwise.

Proof. Recall that for any point $p \in \mathcal{M}$, by (30) we have that $\rho(p \oplus v, p) = \|v\|$ with the Euclidean norm and for all $v \in T_p\mathcal{M}$ within the injectivity radius of $\mathcal{M}$. In other words, the projection of vector v

that lies in the tangent space of base point p has a length of $\|v\|_\infty$. Now, observe that

$$\mu(\boldsymbol{c}^{(J)}, \widetilde{\boldsymbol{c}}^{(J)}) = \mu(\boldsymbol{c}^{(J)}, \mathcal{T}_{\boldsymbol{\alpha}}\boldsymbol{c}^{(J-1)}) + \mu(\mathcal{T}_{\boldsymbol{\alpha}}\boldsymbol{c}^{(J-1)}, \mathcal{T}_{\boldsymbol{\alpha}}\widetilde{\boldsymbol{c}}^{(J-1)}) + \mu(\mathcal{T}_{\boldsymbol{\alpha}}\widetilde{\boldsymbol{c}}^{(J-1)}, \widetilde{\boldsymbol{c}}^{(J)})$$

$$\leq \|\boldsymbol{d}^{(J)}\|_\infty + S_{\mathcal{T}}\mu(\boldsymbol{c}^{(J-1)}, \widetilde{\boldsymbol{c}}^{(J-1)}) + \|\widetilde{\boldsymbol{d}}^{(J)}\|_\infty.$$

Iterating the latter triangle inequality for the middle term gives

$$\mu(\boldsymbol{c}^{(J)}, \widetilde{\boldsymbol{c}}^{(J)}) \leq \|\boldsymbol{d}^{(J)}\|_\infty + \|\boldsymbol{d}^{(J-1)}\|_\infty + S_{\mathcal{T}}^2 \mu(\boldsymbol{c}^{(J-2)}, \widetilde{\boldsymbol{c}}^{(J-2)}) + \|\widetilde{\boldsymbol{d}}^{(J-1)}\|_\infty + \|\widetilde{\boldsymbol{d}}^{(J)}\|_\infty,$$

which inductively yields to

$$\mu(\boldsymbol{c}^{(J)}, \widetilde{\boldsymbol{c}}^{(J)}) \leq \sum_{i=1}^{J} \|\boldsymbol{d}^{(i)}\|_\infty + S_{\mathcal{T}}^J \mu(\boldsymbol{c}^{(0)}, \widetilde{\boldsymbol{c}}^{(0)}) + \sum_{i=1}^{J} \|\widetilde{\boldsymbol{d}}^{(i)}\|_\infty.$$

The required is thus obtained with $L = 1$ if $S_{\mathcal{T}} \leq 1$ and $L = S_{\mathcal{T}}^J$ otherwise.   □

Note that if the two pyramids in Theorem 5.10 were generated by (31) to represent samples of two differentiable curves Γ and $\widetilde{\Gamma}$, respectively. Then, by making use of Corollary 5.9, the sum term in (37) can be bounded in terms of the constants K, P, $\|\nabla\Gamma\|_\infty$ and $\|\nabla\widetilde{\Gamma}\|_\infty$, which depend on the geometry of $\mathcal{M}$.

A special case, where stability in the spirit of Theorem 2.4 can be obtained intrinsically, is when the curvature of the manifold is bounded. Next, we present such a result, assuming $\mathcal{M}$ is complete, open manifold with non-negative sectional curvature. For that we recall two classical theorems: the first and second Rauch comparison theorems (the second is actually due to Berger), tailored to our settings and notation. For more details see Gromoll & Walschap (2009, Chapter 3) and the references therein.

We use the following notation. Denote by $p_k \in \mathcal{M}$ two points, $k = 1, 2$, and $v_k \in T_{p_k}\mathcal{M}$ their vectors in the tangent spaces such that $\|v_1\| = \|v_2\|$, and the value is smaller than the injectivity radius of $\mathcal{M}$. Let $\mathrm{G}(p_1, p_2)$ be the geodesic line connecting p_1 and p_2, and $\mathrm{PG}_{p_2}(v_1) \in T_{p_2}\mathcal{M}$ be the parallel transport of v_1 along $\mathrm{G}(p_1, p_2)$ to $T_{p_2}\mathcal{M}$. Then, the first Rauch theorem suggests that

$$\rho\big(p_2 \oplus v_2, p_2 \oplus \mathrm{PG}_{p_2}(v_1)\big) \leq \|v_2 - \mathrm{PG}_{p_2}(v_1)\|. \tag{38}$$

Moreover, the second Rauch theorem implies that

$$\rho\big(p_1 \oplus v_1, p_2 \oplus \mathrm{PG}_{p_2}(v_1)\big) \leq \rho(p_1, p_2). \tag{39}$$

We are ready for the stability conclusion.

THEOREM 5.11   Let $\mathcal{M}$ be a complete, open manifold with non-negative sectional curvature. Denote by $\{\boldsymbol{c}^{(0)}; \boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(J)}\}$ and $\{\widetilde{\boldsymbol{c}}^{(0)}; \widetilde{\boldsymbol{d}}^{(1)}, \ldots, \widetilde{\boldsymbol{d}}^{(J)}\}$ two pyramids of sequences such that $\|d_k^{(\ell)}\| = \|\widetilde{d}_k^{(\ell)}\|$ and with values smaller than the injectivity radius of $\mathcal{M}$, for all $\ell = 1, \ldots, J$ and $k \in \mathbb{Z}$. Also, assume that $\mu(\boldsymbol{c}^{(0)}, \widetilde{\boldsymbol{c}}^{(0)})$ is sufficiently small so geodesics exist between all pairs $c_k^{(\ell)}, \widetilde{c}_k^{(\ell)}$ for all $\ell = 1, \ldots, J$ and $k \in \mathbb{Z}$, as reconstructed from the above two data pyramids via (32). Assume $\mathcal{T}_{\boldsymbol{\alpha}}$ is stable with constant

$S_{\mathcal{T}}$, as in Definition 5.5. Then, the synthesis sequences $\boldsymbol{c}^{(J)}$ and $\widetilde{\boldsymbol{c}}^{(J)}$ satisfy

$$\mu(\boldsymbol{c}^{(J)}, \widetilde{\boldsymbol{c}}^{(J)}) \leq L\left(\mu(\boldsymbol{c}^{(0)}, \widetilde{\boldsymbol{c}}^{(0)}) + \sum_{i=1}^{J} \|\widehat{\boldsymbol{d}}^{(i)} - \widetilde{\boldsymbol{d}}^{(i)}\|_{\infty}\right), \tag{40}$$

where $\widehat{d}_k^{(i)} = \mathrm{PG}_{(\mathcal{T}_{\boldsymbol{\alpha}} \widetilde{\boldsymbol{c}}^{(i-1)})_k}(d_k^{(i)})$, with $L = 1$ if $S_{\mathcal{T}} \leq 1$ and $L = S_{\mathcal{T}}^J$ otherwise.

Proof. Observe that

$$\mu(\boldsymbol{c}^{(J)}, \widetilde{\boldsymbol{c}}^{(J)}) \leq \mu\left(\boldsymbol{c}^{(J)}, \mathcal{T}_{\boldsymbol{\alpha}} \widetilde{\boldsymbol{c}}^{(J-1)} \oplus \widehat{\boldsymbol{d}}^{(J)}\right) + \mu\left(\mathcal{T}_{\boldsymbol{\alpha}} \widetilde{\boldsymbol{c}}^{(J-1)} \oplus \widehat{\boldsymbol{d}}^{(J)}, \widetilde{\boldsymbol{c}}^{(J)}\right)$$

$$= \mu\left(\mathcal{T}_{\boldsymbol{\alpha}} \boldsymbol{c}^{(J-1)} \oplus \boldsymbol{d}^{(J)}, \mathcal{T}_{\boldsymbol{\alpha}} \widetilde{\boldsymbol{c}}^{(J-1)} \oplus \widehat{\boldsymbol{d}}^{(J)}\right) + \mu\left(\mathcal{T}_{\boldsymbol{\alpha}} \widetilde{\boldsymbol{c}}^{(J-1)} \oplus \widehat{\boldsymbol{d}}^{(J)}, \mathcal{T}_{\boldsymbol{\alpha}} \widetilde{\boldsymbol{c}}^{(J-1)} \oplus \widetilde{\boldsymbol{d}}^{(J)}\right)$$

$$\leq \mu(\mathcal{T}_{\boldsymbol{\alpha}} \boldsymbol{c}^{(J-1)}, \mathcal{T}_{\boldsymbol{\alpha}} \widetilde{\boldsymbol{c}}^{(J-1)}) + \|\widehat{\boldsymbol{d}}^{(J)} - \widetilde{\boldsymbol{d}}^{(J)}\|_{\infty}$$

$$\leq S_{\mathcal{T}} \mu(\boldsymbol{c}^{(J-1)}, \widetilde{\boldsymbol{c}}^{(J-1)}) + \|\widehat{\boldsymbol{d}}^{(J)} - \widetilde{\boldsymbol{d}}^{(J)}\|_{\infty}.$$

For the first inequality we use the triangle inequality, for the second we use (39) and (38). Lastly, we apply the stability of the subdivision scheme. Iterating the latter yields the required bound. $\square$

We present two brief comments on Theorem 5.11. First, bounding the sectional curvature from below with a positive number clearly does not change the conclusion. Still, if the lower bound is negative, such as in hyperbolic manifolds, estimations (38)–(39) do not hold, and more delicate argument is needed. Secondly, we allow the details to differ only by their mutual angle and not magnitude. We may remove this obstacle using a more technical calculation which we omit here for compactness.

REMARK 5.12 Following the methodology of Grohs & Wallner (2012), together with our estimation (35), an analogue of (40) can be achieved based on proximity to the linear counterparts of our operators. This result is more of asymptotic flavor and it carries less information about the constant L. Nevertheless, it holds for more general class of manifolds.

The stability results support the concept of using the inverse transform (32) for different numerical tasks, as we will see in the next section.

6. Numerical examples

In this section we focus on demonstrating our pyramid transform numerically. We begin with an illustration of the bounds from Theorem 3.1 and Theorem 3.3, emphasizing the importance of mask normalization. Then, we show the application of our multiscale transforms over manifold data to the tasks of denoising and anomaly detection. All MATLAB scripts that include the examples of this section are available online at https://github.com/WaelMattar/Manifold-Multiscale-Representations for reproducibility.

6.1 *Comparing the novel linear decimation operators*

In Section 3 we present new methods for truncating the sequence $\boldsymbol{\gamma}$ to obtain a finite mask. Comparing Theorem 3.1 and Theorem 3.3, and in particular, their upper bounds on the norms of the generated detail
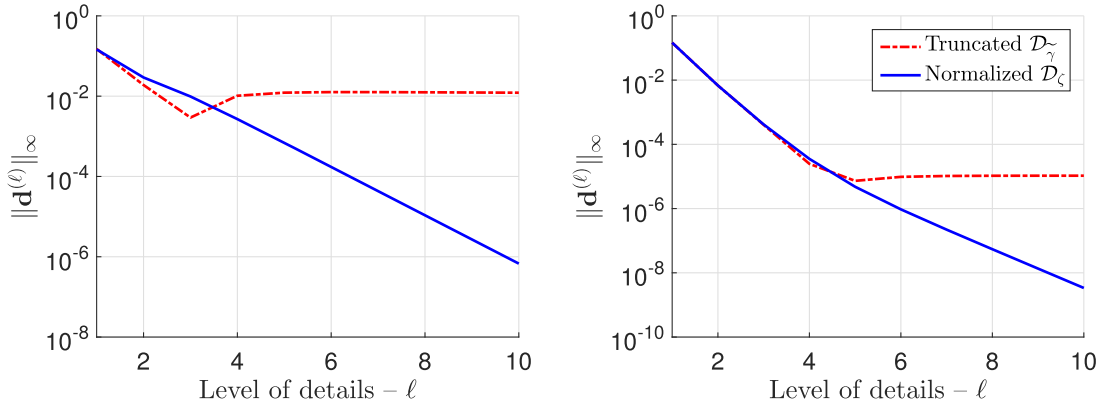
FIG. 3. Truncating the decimation operator: the detail coefficients norm as a function of the level ℓ, plotted on a logarithmic scale for different truncation parameters. On the left, $\varepsilon = 10^{-2}$. On the right, $\varepsilon = 10^{-5}$. In both figures the red dashed lines correspond to $\mathcal{D}_{\widetilde{\gamma}}$ of the truncated mask (17), and the blue lines correspond to $\mathcal{D}_{\zeta}$ of the truncated, shift invariant mask (20). Note how the blue graphs are bounded by geometrically decaying bounds, as guaranteed by Corollary 3.5. In contrast, the red dashed lines are bounded below by values of the same order as the truncation parameter ε, as implied by Theorem 3.1.

coefficients, shows a significant additional factor in (19). This section examines the numerical nature of this difference, and how accurate the description of the detail coefficients' decay according to the theoretical bound is.

Our example is conducted in the functional setting, where we consider the samples of the smooth periodic function $f(x) = \sin(3x)$. We choose $\mathcal{S}_{\boldsymbol{\alpha}}$ to be the linear cubic subdivision scheme, as appears in Example 2.2 and sample f over the interval $[0, 2\pi]$ at 10×2^{10} equispaced points, that is, to obtain $\boldsymbol{c}^{(J)}$ with $J = 10$. The samples are treated as a periodic sequence, so it represents a bi-infinite sequence. Then, we decompose the samples via the linear multiscale transforms (18) and (21), which depend on the truncated mask (17) and shift invariant mask (20), respectively.

The maximum norms of the generated details are depicted in Fig. 3 as a function of the level $\ell = 1, \ldots, 10$, for two different truncation parameters $\varepsilon = 10^{-2}$ and $\varepsilon = 10^{-5}$. The results show behavior that agrees with the upper bounds of Theorem 3.1 and Theorem 3.3. In particular, the details, as generated by (18), are bounded by a value of order ε, due to the additional term in (19), which does not decay with respect to ℓ. On the other hand, the detail coefficients generated by (21) decay geometrically, as expected, see Corollary (3.5).

6.2 *Denoising of sphere-valued curve*

We turn to manifold-valued data and consider the unit sphere $\mathbb{S}^2$ in $\mathbb{R}^3$ as the manifold of this section. The following example serves as a proof of concept for the application of pyramid transform for curves over manifolds. Specifically, we address the problem of estimating a curve from its noisy samples. To this purpose, we follow the conventional algorithm of reconstructing the object from its thresholded multiscale coefficients. For the data model denote by Γ_k, $k \in \mathbb{Z}$ the equidistant samples of a curve Γ over the sphere, and by

$$\Upsilon_k = \Gamma_k \oplus \chi_k, \tag{41}$$

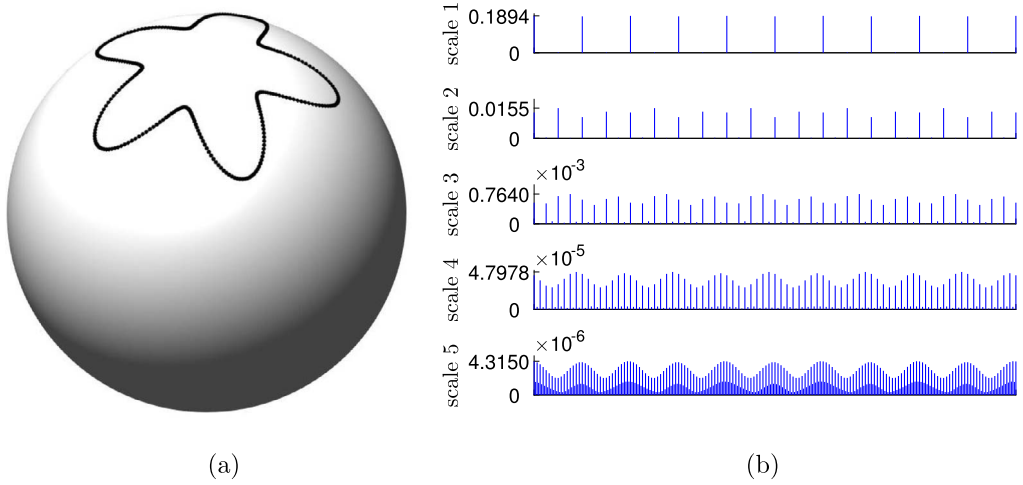Fig. 4. The curve Γ of (42) over the unit sphere with its multiscale transform (31). On the left, 10×2^5 equispaced samples of the flower-like curve. On the right the Euclidean norms of the detail coefficients $d_k^{(\ell)}$ for $\ell = 1, \ldots, 5$, generated by applying (31). As the scale increases the maximal norm of each layer decays geometrically as guaranteed by Corollary 5.9. Note that every second element of each layer is smaller. This phenomenon is explained by condition (11) and reflects the approximation character of our decimation operator. Therefore, smaller values of the truncation parameter ε yield to smaller norms of detail coefficients.

the noisy samples, where $\chi_k \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are i.i.d. normally distributed random variables with zero mean and covariance matrix $\boldsymbol{\Sigma} = \sigma^2 I$. The noise terms χ_k are in the respective tangent spaces $T_{\Gamma_k}\mathcal{M}$, which are isomorphic to $\mathbb{R}^2$. Note that small noise levels guarantee χ_k to be within the injectivity radius of the exponential map $\oplus$ associated to point Γ_k. We, therefore, assume that the realizations of the noise terms are sufficiently small.

In the current test case we take Γ to be a flower-like periodic smooth $\mathbb{S}^2$-valued curve defined via spherical coordinates as,

$$\Gamma(\theta) = \big(\sin(\varphi(\theta))\cos(\theta),\ \sin(\varphi(\theta))\sin(\theta),\ \cos(\varphi(\theta))\big), \quad \varphi(\theta) = \frac{\pi}{16}\cos(N\theta) + \frac{\pi}{6}, \quad \theta \in [0, 2\pi]. \tag{42}$$

Here, $N \in \mathbb{N}$ determines the number of the flower's leaves. We set $N = 5$ as shown in Fig. 4a.

Let $\mathcal{T}_\alpha$ be the Riemannian analogue of the cubic spline subdivision scheme adapted to $\mathbb{S}^2$ as described in Section 2.2. Denote by $\mathcal{Y}_\zeta$ its approximated decimation operator with the shift invariant mask ζ, as given in (31). In this example we pick $\varepsilon = 10^{-5}$, which induces that ζ consists of 13 nonzero elements. We note that $\mathbb{S}^2$ is a two-dimensional topological manifold with positive sectional curvature, thus, optimization problems like (6) and (27) may have infinite solutions, e.g., when averaging two antipodal points. However, for close enough points on $\mathbb{S}^2$, the center of mass exists uniquely, see Dyer *et al.* (2016); Hüning & Wallner (2020). We follow a Riemannian gradient descent method (Krakowski *et al.*, 2007) to calculate the Riemannian center of mass on $\mathbb{S}^2$. Figure 4 demonstrates Γ of (42) alongside its corresponding pyramidical representation via our multiscale transform (31), which manifests the detail coefficients decay.

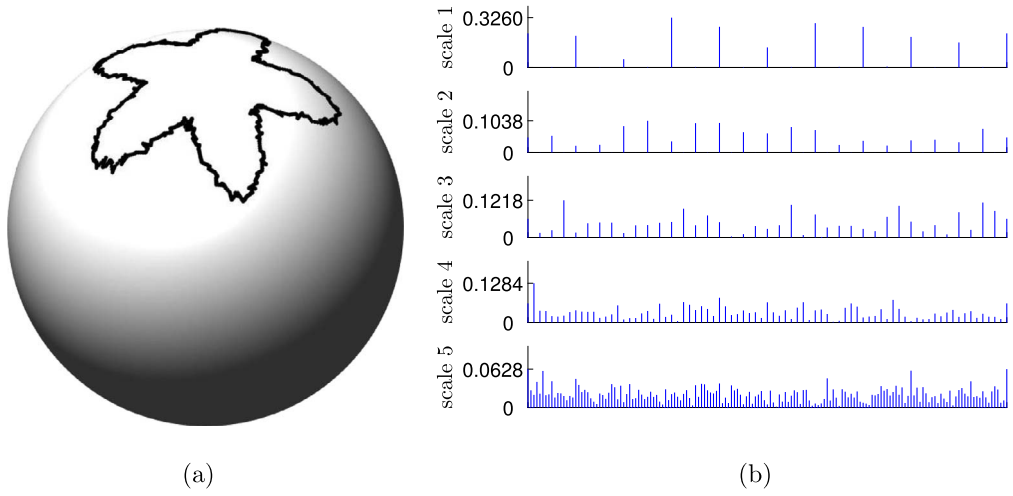(a)   (b)

Fig. 5. Noisy samples and its multiscale representation. On the left the noisy points Υ of (41). On the right the Euclidean norms of the detail coefficients.

We now synthetically generate noisy samples according to the model (41), with $\sigma \approx 1/80$. Figure 5 shows the noisy data alongside their corresponding pyramidical representation via our multiscale transform (31). As we can see, the multiscale representation of the noisy sequence Υ does not enjoy the property of detail coefficients decay.

To estimate Γ from its noisy samples Υ_k we follow Donoho (1995), where it is shown that thresholding of the details of the pyramid transform yields a nearly optimal estimation. In other words, we go over each layer of multiscale coefficients corresponding to the noisy curve, see Fig. 5b, and set to zero all detail coefficients with norm below a fixed threshold, 0.14 in our case. This process yields to a sparser pyramid representation, which forms an estimation of the ground truth Γ. The approximant is synthesized iteratively by (32).

Figure 6 demonstrates the denoised curve alongside its multiscale representation. Indeed, the detail coefficients of the denoised curve are bounded by a geometrically decreasing sequence, which indicates the smoothness of the resulted curve.

To sum our multiscale transform, (31) makes a useful tool for denoising curves over manifolds. The denoising's performance in this example is reflected by the resemblance between the ground truth and the denoised curves.

6.3   *Anomaly detection of $\mathcal{SPD}(3)$-valued curve*

Our multiscale transform (31) involves the application of two local operators. This feature makes the transform a beneficial tool for detecting and analyzing local behavior in manifold-valued curves. This section focuses on representing curves over the cone of 3×3 symmetric positive matrices, which we denote by $\mathcal{SPD}(3)$. In particular, we show the application of our pyramid analysis to the problem of anomaly detection. Namely, we aim to automatically detect rapid local changes in a time series of matrices by inspecting its multiscale representation.

We consider a smooth periodic $\mathcal{SPD}(3)$-valued curve given explicitly via trigonometric deformations. Then, we apply a scaling factor to the eigenvalues of all the matrices that fall in the middle third
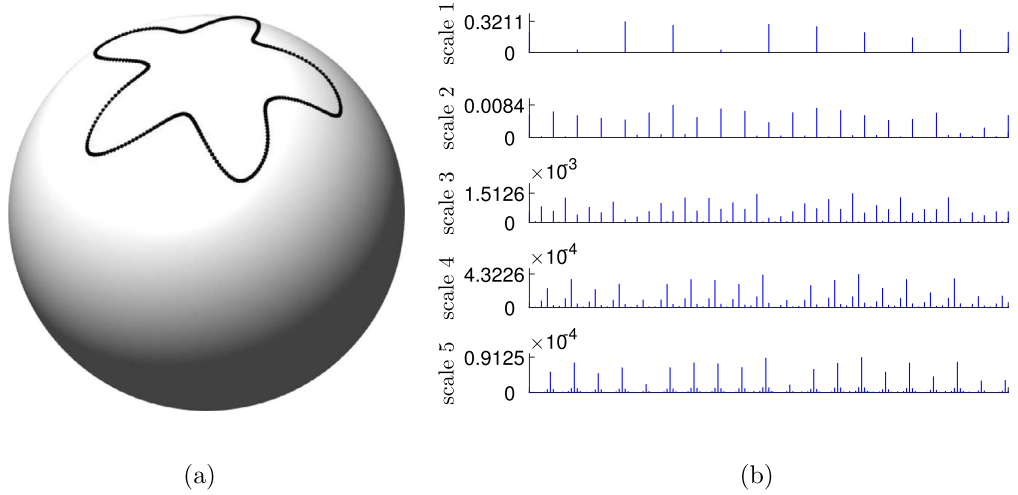
(a) (b)

FIG. 6. Denoised curve and its multiscale representation. On the left estimation of Γ (42). On the right the Euclidean norms of the detail coefficients. The decay of detail coefficients indicate the smoothness of the denoised curve.
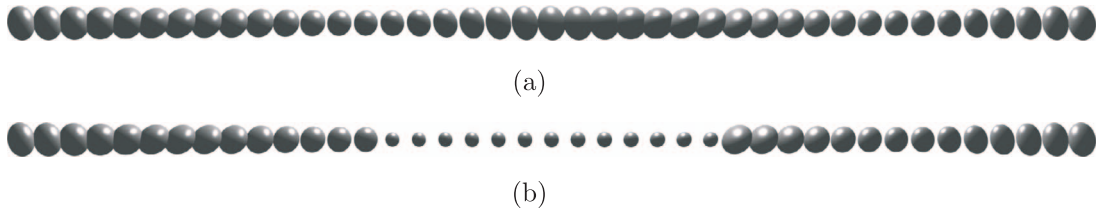

(a)

(b)

FIG. 7. $\mathcal{SPD}(3)$-valued curves. On the top 41 ellipsoids that represent smooth $\mathcal{SPD}(3)$-valued curve. On the bottom the modified 41 ellipsoids now represent the piecewise-smooth curve, with two jump discontinuities in the middle.

of the curve to provide anomaly. This application gives rise to a piecewise smooth $\mathcal{SPD}(3)$-valued curve with two jump discontinuities. We depict the two curves, both the smooth original one and the distributed piecewise smooth, in Fig. 7. Each curve is represented by a series of centered ellipsoids, where every ellipsoid has its main axes determined by the eigenvectors of the corresponding matrix and their lengths by the associated eigenvalues.

We set the test by taking $\mathcal{T}_\alpha$ to be the corner-cutting (quadratic B-spline) subdivision scheme, as presented in Example 2.1, adapted to $\mathcal{SPD}(3)$ as described in Section 2.2. Denote by $\mathcal{Y}_\zeta$ its approximated decimation operator with the truncation parameter $\varepsilon = 10^{-4}$, implying a shift invariant mask ζ with nine nonzeros. The Riemannian center of mass over $\mathcal{SPD}(3)$ is globally unique due to the manifold's nonpositive sectional curvature. To calculate it we follow the gradient descent method in Iannazzo *et al.* (2019).

Next, we decompose both curves of Fig. 7 by the multiscale transform (31) and investigate the norms of the detail coefficients. The norms of the detail coefficients, which lie in the linear space of all symmetric matrices of order 3, are presented in Fig. 8. As it turns out, the detail coefficients corresponding to the smooth curve are represented by a geometrically decreasing sequence, as guaranteed by Corollary 5.9. However, in the vicinities of the anomaly points, the detail coefficients
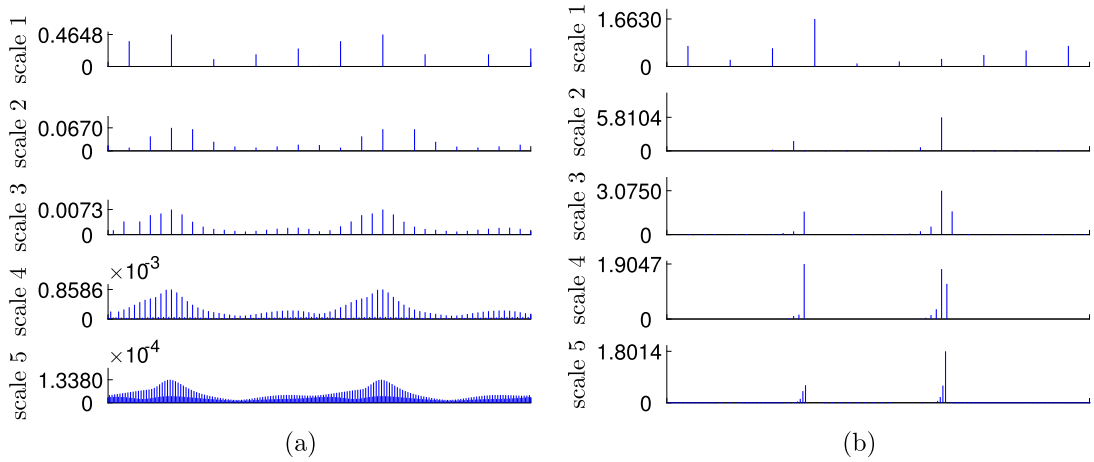
FIG. 8. Frobenius norms of the detail coefficients of the two $\mathcal{SPD}(3)$ curves in Fig. 7. On the left detail coefficients norms corresponding to Fig. 7a. On the right detail coefficients norms corresponding to Fig. 7b. The decay rate, which decreases with each layer in (a), implies the curve's smoothness. Moreover, note how the theoretical condition (11) is illustrated in (a) as every second detail is proportional to the truncation parameter ε. On the other hand, the two local peaks in (b) indicate radical changes in the respective curve and reveal the abnormalities.

generated by our multiscale transform (31) have relatively large norms. Namely, the large detail coefficients are correlated with the parametric locations around the jump discontinuities. Therefore, the multiscale transform (31) makes a useful tool for detecting such anomalies.

REMARK 6.1 We numerically estimated the constant P of (35) corresponding to this section's manifold settings. The results appear in Table A1 and Table A2 in Appendix A where we present the minimal possible P. This value decreases monotonically to 1 as the scale of sampling, J, increases. This phenomenon implies that the decimation operation $\mathcal{Y}_\zeta$ behaves like the simple downsampling operation for close enough $\mathcal{M}$-valued data points.

Acknowledgements

Funding

REFERENCES

BARBARESCO, F. & GAY-BALMAZ, F. (2020) Lie group cohomology and (multi) symplectic integrators: new geometric tools for Lie group machine learning based on souriau geometric statistical mechanics. *Entropy*, **22**, 498.

BLANES, S. & CASAS, F. (2017) *A Concise Introduction to Geometric Numerical Integration*. Boca Raton, FL: Chapman and Hall/CRC.

DO CARMO, M. P. (1992) *Riemannian Geometry*. Boston: Birkhäuser.

DAUBECHIES, I. (1992) *Ten Lectures on Wavelets*. Philadelphia, Pennsylvania: SIAM.

DE BOOR, C. (1978) *A Practical Guide to Splines*, vol. **27**. New York: Springer.

DONOHO, D. L. (1992) Interpolating Wavelet Transforms. Preprint, Department of Statistics, Stanford University, **2**, 1–54.

DONOHO, D. L. (1995) De-noising by soft-thresholding. *IEEE Trans. Inf. Theory*, **41**, 613–627.

DYER, R., VEGTER, G., & WINTRAECKEN, M. (2016) *Barycentric coordinate neighbourhoods in Riemannian manifolds*. arXiv preprint arXiv:1606.01585.

DYN, N. (1992) *Subdivision Schemes in Computer-Aided Geometric Design. Advances in Numerical Analysis, II, Wavelets, Subdivision Algorithms and Radial Basis Functions*, vol. **621**. Oxford, i992. ll.: Clarendon Press, pp. 36–104.

DYN, N. (2002a) Analysis of convergence and smoothness by the formalism of Laurent polynomials. *Tutorials on Multiresolution in Geometric Modelling*. Berlin, Heidelberg: Springer, pp. 51–68.

DYN, N. (2002b) Interpolatory subdivision schemes. *Tutorials on Multiresolution in Geometric Modelling*. Berlin, Heidelberg: Springer, pp. 25–50.

DYN, N. & SHARON, N. (2017a) A global approach to the refinement of manifold data. *Math. Comp.*, **86**, 375–395.

DYN, N. & SHARON, N. (2017b) Manifold-valued subdivision schemes based on geodesic inductive averaging. *J. Comput. Appl. Math.*, **311**, 54–67.

DYN, N. & ZHUANG, X. (2020) Linear multiscale transforms based on even-reversible subdivision operators. *Excursions in Harmonic Analysis, Volume 6: In Honor of John Benedetto's 80th Birthday*, vol. **6**. Boston, MA: Springer.

FRANK, J. & OURMAZD, A. (2016) Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods*, **100**, 61–67.

GRÖCHENIG, K. (2010) Wiener's lemma: theme and variations. an introduction to spectral invariance and its applications. *Four Short Courses on Harmonic Analysis: Wavelets, Frames, Time-Frequency Methods, and Applications to Signal and Image Analysis*. Boston: Birkhäuser, pp. 175–234.

GROHS, P. (2010a) A general proximity analysis of nonlinear subdivision schemes. *SIAM J. Math. Anal.*, **42**, 729–750.

GROHS, P. (2010b) Stability of manifold-valued subdivision schemes and multiscale transformations. *Constr. Approx.*, **32**, 569–596.

GROHS, P. & WALLNER, J. (2009) Interpolatory wavelets for manifold-valued data. *Appl. Comput. Harmon. Anal.*, **27**, 325–333.

GROHS, P. & WALLNER, J. (2012) Definability and stability of multiscale decompositions for manifold-valued data. *J. Franklin Inst.*, **349**, 1648–1664.

GROMOLL, D. & WALSCHAP, G. (2009) *Metric Foliations and Curvature*, vol. **268**. Berlin, Germany: Springer Science & Business Media.

GROVE, K. & KARCHER, H. (1973) How to conjugatec 1-close group actions. *Math. Z.*, **132**, 11–20.

HARDERING, H. (2015) Intrinsic discretization error bounds for geodesic finite elements. *Ph.D. Thesis*, Berlin: FU.

HARTEN, A. (1996) Multiresolution representation of data: a general framework. *SIAM J. Numer. Anal.*, **33**, 1205–1256.

HÜNING, S. & WALLNER, J. (2020) Convergence analysis of subdivision processes on the sphere. *IMA J. Numer. Anal.*, **00**, 1–14.

IANNAZZO, B., JEURIS, B. & POMPILI, F. (2019) The derivative of the matrix geometric mean with an application to the nonnegative decomposition of tensor grids. *Structured Matrices in Numerical Linear Algebra: Analysis, Algorithms and Applications*. Cham, Switzerland: Springer, pp. 107–128.

ISERLES, A., MUNTHE-KAAS, H. Z., NØRSETT, S. P. & ZANNA, A. (2000) Lie-group methods. *Acta Numer.*, **9**, 215–365.

KARCHER, H. (1977) Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, **30**, 509–541.

KARCHER, H. (2014) Riemannian center of mass and so called Karcher mean. arXiv preprint arXiv:1407.2087.

KAUR, A. & SINGH, C. (2017) Contrast enhancement for cephalometric images using wavelet-based modified adaptive histogram equalization. *Appl. Soft Comput.*, **51**, 180–191.

KOBAYASHI, S. & NOMIZU, K. (1963) *Foundations of Differential Geometry*, vol. **1**. New York: John Wiley & Sons Inc.

KRAKOWSKI, K., HÜPER, K. & MANTON, J. (2007) On the computation of the Karcher mean on spheres and special orthogonal groups. *Conference Paper, Robomat*. Coimbra, Portugal: Citeseer.

LANE, J. M. & RIESENFELD, R. F. (1980) A theoretical development for the computer generation and display of piecewise polynomial surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, **PAMI-2**, 35–46.

LUNGA, D., PRASAD, S., CRAWFORD, M. M. & ERSOY, O. (2014) Manifold-learning-based feature extraction for classification of hyperspectral data: a review of advances in manifold learning. *IEEE Signal Process. Mag.*, **31**, 55–66.

LV, X., LIAO, X. & YANG, B. (2018) A novel scheme for simultaneous image compression and encryption based on wavelet packet transform and multi-chaotic systems. *Multimed. Tools Appl.*, **77**, 28633–28663.

MALLAT, S. (1999) *A Wavelet Tour of Signal Processing*. Burlington, MA: Elsevier.

RAHMAN, I. U., DRORI, I., STODDEN, V. C., DONOHO, D. L. & SCHRÖDER, P. (2005) Multiscale representations for manifold-valued data. *Multiscale Model. Simul.*, **4**, 1201–1232.

SANDER, O. (2016) Geodesic finite elements of higher order. *IMA J. Numer. Anal.*, **36**, 238–266.

STORATH, M. & WEINMANN, A. (2020) Wavelet sparse regularization for manifold-valued data. *Multiscale Model. Simul.*, **18**, 674–706.

STROHMER, T. (2002) Four short stories about toeplitz matrix calculations. *Linear Algebra Appl.*, **343**, 321–344.

WALLNER, J. (2020) Geometric subdivision and multiscale transforms. *Handbook of Variational Methods for Nonlinear Geometric Data*. Cham, Switzerland: Springer, pp. 121–152.

WALLNER, J. & DYN, N. (2005) Convergence and analysis of subdivision schemes on manifolds by proximity. *Comput. Aided Geom. Des.*, **22**, 593–622.

ZEILMANN, A., SAVARINO, F., PETRA, S. & SCHNÖRR, C. (2020) Geometric numerical integration of the assignment flow. *Inverse Probl.*, **36**:034003.

A. Numerical evaluation of the decaying factor

Lemma 5.8 introduces a decaying rate of the norms of the details. Here, we provide several numerical evaluations of the decaying factor P of (35), as observed in the examples of Sections 6.2 and 6.3. Indeed, there exists a constant $P > 1$ such that (35) holds for sequences c sampled equidistantly from a differentiable curve over arc-length parametrization. In particular, as seen through the proof of Lemma 5.8, the minimal possible value of P can be evaluated by

$$P_{\min} = \frac{\Delta_{\mathcal{M}}(\mathcal{Y}_\xi c)}{2\Delta_{\mathcal{M}}(c)},$$

where $\mathcal{Y}_\xi$ is the decimation operator used in the multiscale transform 31. Under the settings of Sections 6.2 and 6.3 we calculate $P_{\min}$ for different $\Delta_{\mathcal{M}}(c)$ values. The results are shown in Table A1 and Table A2.

The main feature of Table A1 and Table A2 is that, in both manifold settings, the constant $P_{\min}$ decreases monotonically to 1 as $\Delta_{\mathcal{M}}(c)$ decreases. This fact indicates a similar behavior between $\mathcal{Y}_\xi$ and the downsampling operation $\downarrow 2$, when the distance between data points reduces, as stated in Remark 6.1.

TABLE A1 *The constant P_{min} against different values of $\Delta_{\mathcal{M}}(c)$. The second row demonstrates P_{min} corresponding to samples of the curve Γ in (42) over the sphere $\mathbb{S}^2$, with the respective $\Delta_{\mathcal{M}}(c)$ value. The operator $\mathcal{Y}_\xi$ is the adapted even-inverse of the cubic B-spline with truncation parameter $\varepsilon = 10^{-5}$, as in Section 6.2*

$\Delta_{\mathcal{M}}(c)$	0.2667	0.1639	0.0859	0.0433	0.0217	0.0108	0.0054	0.0027
$\mathbb{S}^2$	1.4021	1.0368	1.0205	1.0086	1.0038	1.0003	1.0001	1.0000

TABLE A2 *The constant P_{min} against different values of $\Delta_{\mathcal{M}}(c)$. The second row demonstrates P_{min} corresponding to samples of the curve shown in Fig. 7a over the manifold $\mathcal{SPD}(3)$, with the respective $\Delta_{\mathcal{M}}(c)$ value. The operator $\mathcal{Y}_\xi$ is the adapted even-inverse of the quadratic B-spline with truncation parameter $\varepsilon = 10^{-4}$, as in Section 6.3*

$\Delta_{\mathcal{M}}(c)$	0.6837	0.3542	0.1813	0.0912	0.0457	0.0228	0.0114	0.0057
$\mathcal{SPD}(3)$	1.2661	1.0613	1.0176	1.0053	1.0014	1.0004	1.0000	1.0000