IMA Journal of Numerical Analysis (2013) **33**, 849–874
doi:10.1093/imanum/drs026
Advance Access publication on October 31, 2012

Quasi-interpolation in Riemannian manifolds

PHILIPP GROHS
Seminar for Applied Mathematics, ETH Zürich, Rämistrasse 101, 8004 Zürich Switzerland
pgrohs@math.ethz.ch, philippgrohs@gmail.com

[Received on 28 January 2012; revised on 21 January 2012]

We consider quasi-interpolation operators for functions assuming their values in a Riemannian manifold. We construct such operators from corresponding linear quasi-interpolation operators by replacing affine averages with the Riemannian centre of mass. As a main result, we show that the approximation rate of such a nonlinear operator is the same as for the linear operator it has been derived from. In order to formulate this result in an intrinsic way, we use the Sasaki metric to compare the derivatives of the function to be approximated with the derivatives of the nonlinear approximant. Numerical experiments confirm our theoretical findings.

Keywords: Quasiinterpolation, Riemannian Data, Geodesic Finite Elements, Approximation Order, Riemannian Center of Mass.

1. Introduction

A fundamental problem in computational science is the handling of massive amounts of data. In addition to the sheer mass of data to be processed, in recent times many modern sensing mechanisms produce data which is of a nonstandard type, with data points assuming their value in nonlinear geometries. Examples include the following subjects.

(1) Deformation tensors, where the data points consist of elements of the Cartan–Hadamard space of positive-definite symmetric matrices. The data are modelled as a function $\mathbb{R}^3 \to \mathrm{SPD}(3)$. Such data arise, for instance, in diffusion tensor magnetic resonance imaging (MRI) in medical imaging (Denis Le Bihan *et al.*, 2001) or strain and stress measurement in materials science.

(2) Positions of rigid bodies, where the data points consist of elements of the Lie group of rigid body motions. The data are modelled as a function $\mathbb{R} \to \mathrm{SE}(3)$. Data of this type arise, for instance, in kinematics or motion design (Wallner & Pottmann, 2006), or Cosserat rod modelling (Sander, 2010).

(3) Orientations, where the data points consist of elements of the Lie group of orthogonal matrices. The data are modelled as a function $\mathbb{R} \to \mathrm{O}(3)$. Orientation-valued data arrays arise, for instance, as 'black-box' recordings of the orientation of aircraft, varying with time (Rahman *et al.*, 2006).

(4) Subspaces, where the data points consist of elements of a Grassmanian manifold. The data are modelled as a function $\mathbb{R} \to \mathrm{G}(k,n)$. These data types can arise, for instance, in array signal processing (Rahman *et al.*, 2006).

(5) Orthogonal matrices with positive determinants, where data are modelled as a function $\mathbb{R} \to \mathrm{SO}(n)$. Such data arise, e.g, in isospectral flow problems (Iserles *et al.*, 2000).

This (incomplete) list suggests that it is of eminent interest to develop useful computational and theoretical tools capable of processing manifold-valued functions. In this spirit, the objective of this paper is to present a complete extension of the theory of quasi-interpolation (de Boor & Fix, 1973; Chui, 1992) to the nonlinear case. We will start with a linear quasi-interpolation scheme and, by replacing affine averages with the Riemannian centre of mass (Karcher, 1977), wind up with an intrinsic approximation procedure for manifold-valued functions. Our main result is that the linear properties completely carry over to the nonlinear case.

Even stating such a result is nontrivial since it is at first glance not clear how to compare a differential of a function with the differential of an approximant, both of which assume their values in the (iterated) tangent bundle of the manifold. We solve this problem by utilizing the so-called Sasaki metric (Sasaki, 1958) which is a canonical Riemannian structure defined on the tangent bundle of a Riemannian manifold. Using this formulation, we are able to give a complete and intrinsic extension of the linear theory; see Theorem 3.8.

To give a flavour of our main result consider a linear quasi-interpolation operator

$$f \mapsto \bar{Q}^h f(\cdot) := \sum_{j\in\mathbb{Z}} f(hj)\Phi(h^{-1}\cdot - j),$$

where Φ might be, for instance, an affine combination of the integer translates of the fundamental cardinal cubic B-spline function as in Example 2.5 (of course, higher orders than cubic and different functions are possible).

The well-known linear theory establishes results regarding the decay of the approximation error,

$$\left\| \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^l (f - \bar{Q}^h f) \right\|_\infty ,$$

with the step width h tending to zero. This approximation error in h is related to a few simple properties of Φ such as smoothness and polynomial exactness; see Section 2.1.

Our main idea is to regard the expression for $\bar{Q}^h f$ as a weighted average of the samples $f(hj)$. By replacing affine averages with the Riemannian centre of mass (Karcher, 1977), we arrive at a definition of an associated quasi-interpolation operator $Q^h f$ for functions f with values in a differentiable manifold $\mathcal{M}$; see Section 2.2.

In order to study the approximation errors of Q^h, we need to be able to compare derivatives $d^l f$ with $d^l Q^h f$, both taking their values in the iterated tangent bundle $T^l\mathcal{M}$ of $\mathcal{M}$. This is achieved by using the Sasaki metric on $T^l\mathcal{M}$ which is described in Section 3.2. With the geodesic distance $\mathfrak{s}_l$ induced by the Sasaki metric, our main result is that the approximation error

$$\mathfrak{s}_l(d^l f, d^l Q^h f)$$

behaves in the exact same way as the corresponding linear error; see Theorem 3.8 for a more precise statement.

Even though this result might not be very surprising, its proof turns out to be quite involved.

As an application, we construct nonlinear approximation manifolds (see Definition 2.7) with a prescribed number of degrees of freedom and determine their approximation properties.

Besides approximating an explicitly given function, a main motivation for our construction is the potential use of these approximation manifolds to solve manifold-valued optimization problems in a finite-element-like fashion (cf. Sander, 2010, 2011).

We would like to note that our construction can easily be adapted to the multivariate case and also over simplicial grids (cf. Sander, 2011). The study of the associated approximation properties (also with respect to other error norms) will be the subject of forthcoming work.

1.1 *Previous work*

There exists, by now, a substantial body of previous work related to nonlinear data types of which we only mention some examples. In the paper of Rahman *et al.* (2006), a manifold-valued wavelet transform has been derived, and its theoretical properties are investigated in Grohs & Wallner (2009), Grohs (2010c) and Harizanov & Oswald (2010). The idea to use nonlinear subdivision schemes for the approximation of manifold-valued data was investigated in Grohs (2010a,b), Weinmann (2010), Xie & Yu (2008, 2010, 2011), Wallner & Dyn (2005), Wallner & Pottmann (2006). Among these, we would like to single out the work of Xie & Yu (2011) where a different construction of quasi-interpolants is presented and analysed numerically. The already-mentioned papers (Sander, 2010, 2011) present a construction of first-order geodesic finite element spaces much in the spirit of our (higher-order) construction. Related to these constructions, various optimization problems are studied. Finally, we would like to mention (Iserles *et al.*, 2000), where Runge–Kutta methods are extended to the case of Lie-valued (and more general) ordinary differential equations (ODEs). On the theoretical side, Karcher (1977) introduced a general mollifying procedure for the approximation of functions between Riemannian manifolds. One can interpret our method as a specimen of his construction. In contrast to Karcher (1977), we are able to derive the exact same approximation rates as corresponding linear constructions for our construction, even for the approximation of higher-order derivatives.

1.2 *Outline*

The outline is as follows. In Section 2, we will lay out the necessary background from linear approximation theory and also describe the nonlinear set-up we will work with. After that, in Section 3, we will first show that it suffices to study our approximation problems in a chart, and then we will discuss the Sasaki metric which will allow us to formulate our results in an intrinsic fashion. Section 4 contains the proof of our main result and forms the main technical part of this paper. Finally, in Section 5, we present some numerical computations which confirm our theory. We also include an appendix containing some auxiliary results which will be needed in the course of the proof and another one explaining some basic concepts in Riemannian geometry which will be needed in the course of this paper.

1.3 *Notation*

For a function $f : \mathbb{R} \to \mathbb{R}^d$, we use the usual terminology $\|f\|_\infty := \sup_x |f(x)|$ with $|\cdot|$ the maximum norm on $\mathbb{R}^d$. The space of continuous functions with values in a manifold $\mathcal{M}$ is denoted by $C^0(\mathbb{R}, \mathcal{M})$. If $\mathcal{M}$ is a Euclidean space, we simply write C^0. Furthermore, we use the symbol l_∞ for the Banach space of bounded, real-valued sequences on $\mathbb{Z}$. We use boldface notation for multi-indices $\mathbf{j} \in \mathbb{Z}^k$ and denote $|\mathbf{j}|_1 := j_1 + \cdots + j_k$. Distance metrics on Riemannian manifolds are usually expressed by fraktur letters, for instance, $\mathfrak{d}$. The space of polynomials of degree $\leqslant m$ on $\mathbb{R}$ shall be denoted by Π_m. For a set A, the symbol χ_A denotes its indicator function.

2. Preliminaries

2.1 *Linear theory*

We start by reviewing some well-known facts from linear quasi-interpolation theory. For more information, we refer to de Boor & Fix (1973), Chui (1992) and DeVore & Lorentz (1993). In the following definition, we introduce the smoothness spaces C^α which we will work with.

DEFINITION 2.1 For a function $f:\mathbb{R}\to\mathbb{R}^d$, we define $\Delta_h^l f$, the lth-order forward difference with step width $h>0$ inductively by

$$\Delta_h^0 f := f \quad \text{and} \quad \Delta_h^l f(x) := \Delta_h^{l-1}(x+h) - \Delta_h^{l-1}(x).$$

We say that $f\in C^\alpha$, $\alpha>0$ if we have

$$\left\| \Delta_h \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^{\lfloor\alpha\rfloor} f \right\|_\infty = \mathcal{O}(h^{\alpha-\lfloor\alpha\rfloor}).$$

Linear finite element spaces on a regular partition of an interval are usually constructed from a function $\Phi:\mathbb{R}\to\mathbb{R}$ satisfying the following requirements:

$$\sum_{j\in\mathbb{Z}} \Phi(\cdot - j) = 1 \quad \text{(partition of unity)}, \tag{2.1}$$

$$\exists N\in\mathbb{R}:\ \operatorname{supp}\Phi \subset [-N,N] \quad \text{(locality)}, \tag{2.2}$$

$$\sum_{j\in\mathbb{Z}} p(i)\Phi(\cdot - j) = p(\cdot) \quad \forall p\in\Pi_{m-1} \quad \text{(polynomial exactness)}, \tag{2.3}$$

$$\Phi\in C^s \quad \text{(smoothness)}. \tag{2.4}$$

Given such a function, we can define the *linear finite element spaces*

$$V_h(\Phi) := \left\{ \sum_{j\in\mathbb{Z}} a(j)\Phi(h^{-1}\cdot - j):\ (a(j))_{j\in\mathbb{Z}}\in l_\infty \right\}.$$

In many applications, it is necessary to approximate a given function f by the finite element spaces $V_h(\Phi)$. This can be done either implicitly, when f is given as a solution of an operator equation, or explicitly, when f is explicitly given. One of the very fundamental results in approximation theory is that the approximation rate of the finite element spaces for a given function f is exactly governed by the smoothness of f.

THEOREM 2.2 (Jackson theorem) Assume that $f\in C^\alpha$ with $\alpha<m$. Then for $l<\alpha$ and $l<s$, we have

$$\inf_{g\in V_h(\Phi)} \left\| \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^l (f-g) \right\|_\infty = \mathcal{O}(h^{\alpha-l}). \tag{2.5}$$

In fact, more can be said: under the assumptions (2.1)–(2.4), a quasioptimal approximant $g\in V_h(\Phi)$ can be constructed explicitly.

DEFINITION 2.3 For Φ satisfying (2.1)–(2.4) we define the *linear quasi-interpolation operator* $\bar{Q}^h: C^0\to V_h(\Phi)$ defined via

$$f(\cdot)\mapsto \bar{Q}^h f(\cdot) := \sum_{j\in\mathbb{Z}} f(hj)\Phi(h^{-1}\cdot - j)\in V_h(\Phi). \tag{2.6}$$

Now we can state the following stronger form of Theorem 2.2.

THEOREM 2.4 Assume that $f \in C^\alpha$ with $\alpha < m$. Then, for $l < \alpha$ and $l < s$, we have

$$\left\| \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^l (f - \bar{Q}^h f) \right\|_\infty = \mathcal{O}(h^{\alpha - l}). \tag{2.7}$$

EXAMPLE 2.5 (Cubic B-spline quasi-interpolation) A classical example of a quasi-interpolation operator can be constructed from the cardinal cubic B-spline function $B_3(\cdot)$ (Chui, 1992). Even though the choice $\Phi = B_3$ would only lead to the low-degree polynomial reproduction of 1, it is possible to preprocess the sampling data of f and arrive at a fourth-order quasi-interpolation scheme

$$\bar{Q}^h f(\cdot) := \sum_{j \in \mathbb{Z}} B_3(h^{-1} \cdot - j) \left(-\frac{1}{6} f(h(j-1)) + \frac{4}{3} f(hj) - \frac{1}{6} f(h(j+1)) \right),$$

which falls into our definition by putting

$$\Phi(\cdot) = -\frac{1}{6} B_3(\cdot - 1) + \frac{4}{3} B_3(\cdot) - \frac{1}{6} B_3(\cdot + 1)$$

which can be shown to satisfy (2.3) with $m = 4$ (Chui, 1992). The linear approximation spaces are given by piecewise cubic polynomials, C^2 at the knots.

Clearly, Theorem 2.4 implies Theorem 2.2; for a proof of Theorem 2.4, we refer to Chui (1992). We would like to close this section by noting that the backbone of any approximation method or finite element solver is given by the validity of a Jackson theorem (Braess, 2007).

2.2 *Nonlinear approximation*

Having presented a brief introduction to the construction of regular finite element spaces for functions $f : \mathbb{R} \to \mathbb{R}^d$, we would now like to ask whether such a construction can be meaningfully extended to the case of manifold-valued functions $f : \mathbb{R} \to \mathcal{M}$, $\mathcal{M}$ being a d-dimensional differentiable manifold. More specifically, our goal is to construct a quasi-interpolation operator operating on $\mathcal{M}$-valued functions such that an equivalent to Theorem 2.4 holds. The approach we will take is to regard the sum

$$\sum_{j \in \mathbb{Z}} f(hj) \Phi(h^{-1} \cdot - j)$$

in (2.6) as a weighted average of the points $(f(hj))_{j \in \mathbb{Z}}$ with weights given by $(\Phi(h^{-1} \cdot - j))_{j \in \mathbb{Z}}$. By (2.1) this is justified. The key insight is that weighted averages also exist in Riemannian manifolds, i.e., a pair $(\mathcal{M}, g)$ with $\mathcal{M}$ a differentiable manifold and g a metric tensor field on $\mathcal{M}$; see DoCarmo (1992) (or Appendix B) for more information on Riemannian geometry.

DEFINITION 2.6 Assume that $(\mathcal{M}, g)$ is a Riemannian manifold with induced metric $\mathfrak{d}$. For points $(p(j))_{j \in J}$ and weights $(w(j))_{j \in J}$, the Riemannian centre of mass

$$x^* = \mathrm{av}_{\mathcal{M}}((p(j))_{j \in J}, (w(j))_{j \in J})$$

is defined as

$$x^* = \underset{x \in \mathcal{M}}{\mathrm{argmin}} \sum_{j \in J} w(j) \mathfrak{d}(p(j), x)^2.$$

It can be shown that locally the Riemannian centre of mass exists and is unique. Furthermore, locally, it is characterized by the first-order equilibrium condition

$$\sum_{j\in J} w(j)\log_{\mathcal{M}}(x^*, p(j)) = 0; \tag{2.8}$$

see e.g., Karcher (1977). Here, $\log_{\mathcal{M}}$ denotes the logarithm mapping of $\mathcal{M}$, e.g., the local inverse of the exponential function $\exp_{\mathcal{M}}$ of $\mathcal{M}$ (DoCarmo, 1992).

The natural idea is now to replace the affine weighted average in (2.6) with the Riemannian average in $\mathcal{M}$. This leads to the following definition.

Definition 2.7 We define the *nonlinear finite element manifolds*

$$V_h^{\mathcal{M}}(\Phi) := \{\mathrm{av}_{\mathcal{M}}((a(j))_{j\in\mathbb{Z}}, (\Phi(h^{-1}\cdot - j))_{j\in\mathbb{Z}}) : \text{ whenever the average is well defined}\}.$$

It is our interest to study the approximation properties of these nonlinear finite element manifolds, i.e., instead of approximating with linear spaces we are approximating with nonlinear manifolds with the same number of degrees of freedom. We will do this by considering explicit projection operators onto these manifolds in terms of nonlinear quasi-interpolation as defined below.

Definition 2.8 Define the *nonlinear quasi-interpolation operator*

$$Q^h : C^0(\mathbb{R}, \mathcal{M}) \to C^0(\mathbb{R}, \mathcal{M})$$

by

$$Q^h f(\cdot) := \mathrm{av}_{\mathcal{M}}((f(hj))_{j\in\mathbb{Z}}, (\Phi(h^{-1}\cdot - j)_{j\in\mathbb{Z}}). \tag{2.9}$$

Remark 2.9 Note that for h sufficiently small and $f \in C^\alpha$ for any $\alpha > 0$, the expression $Q^h f$ is always defined. This is due to the locality of Φ and the fact that all sampling points used in the computation of $Q^h f(x)$ lie in a set of arbitrarily small diameter. Therefore, by local well-definedness of the Riemannian centre of mass, the average leading to $Q^h f(x)$ exists. For this reason, we will ignore the issue of well-definedness in the sequel and tacitly assume that the sampling width h is sufficiently small.

We would also like to remark that for several examples of practical interest, the manifold $\mathcal{M}$ possesses particular structural properties which make the Riemannian averages defined for *any* initial data. One such example is the manifold SPD(n) of symmetric positive-definite $n \times n$ matrices arising, e.g., in diffusion tensor MRI in medical imaging (Pennec *et al.*, 2006).

Remark 2.10 Karcher (1977) constructs a similar, continuous mollifying procedure for approximating manifold-valued functions (which may also be defined on a general manifold). The fact that Karcher did not succeed in proving approximation results for higher-order derivatives accounts for the nontrivial nature of our problem. In Karcher (1977, p. 521), he writes

> The approximation of higher derivatives of f, if they exist, by the corresponding derivatives of [*the approximant*] is not clear to me.

We believe that our approach could be suitable for clarifying this question in the general case.

The idea to replace affine averages by the Riemannian centre of mass is not new. In Grohs (2010b) and Wallner *et al.* (2011), it was applied to study smoothness and approximation properties of manifold-valued subdivision schemes. The very recent and interesting work of Sander (2010, 2011) constructs first-order finite element spaces essentially in the same way as we do. These spaces are then used to solve manifold-valued optimization problems arising, e.g., in Cosserat rod modelling in a very natural fashion. We expect our present results to be relevant in this direction. Finally, we would like to mention Pennec *et al.* (2006), where Riemannian methods are used for the filtering, denoising and statistical analysis of diffusion tensor MRI data.

The central question to ask is whether the approximation properties which are known in the linear case (e.g., Theorem 2.4) also hold for the nonlinear quasi-interpolation operators. Answering this question affirmatively will be the main theme of this paper.

In order to pose the question of approximability of the derivatives of f correctly, it is necessary to give a canonical notion of difference between elements of the tangent bundles $T^l\mathcal{M}$. In Section 3, we address this problem and present an appropriate construction, namely, the so-called Sasaki metric (Sasaki, 1958). Then, in Section 4, we show that indeed the linear approximation results also hold in the same form for manifold-valued quasi-interpolation.

3. Localization and natural metrics

In this section, we will examine how we can perform our local computations in a chart without losing any generality. In order to help the reader with little experience in differential geometry, we have added a short Appendix B where most of the geometric terms are explained in a leisurely manner.

3.1 *Computations in charts*

We now consider a chart $\gamma : \mathcal{M} \to \mathbb{R}^d$ and its induced chart $(\gamma, d\gamma) : T\mathcal{M} \to \mathbb{R}^{2d}$. The starting point is the balance equation (2.8), which can be transferred to a similar equation in $\mathbb{R}^d$ by composition with the chart γ.

More precisely, using the notation

$$u(x,y) := \gamma \circ \exp_{\mathcal{M}}(\gamma^{-1}(x), (d\gamma)^{-1}|_x y) \quad \text{and} \quad v(x,y) := d\gamma|_{\gamma^{-1}x}(\log_{\mathcal{M}}(\gamma^{-1}x, \gamma^{-1}y)), \quad x, y \in \mathbb{R}^d,$$

we can write (2.8) as

$$\gamma \circ Q^h f(x) = u\left(\gamma \circ Q^h f(x), \sum_{j\in\mathbb{Z}} \Phi(h^{-1}x - j) v(\gamma \circ Q^h f(x), \gamma \circ f(hj))\right). \tag{3.1}$$

In order to verify (3.1), observe that a point (x, y) can be transported back to the tangent vector $d\gamma^{-1}|_x y \in T_{\gamma^{-1}(x)}\mathcal{M}$ and that the mapping $d\gamma^{-1}|_x$ is linear, cf. Appendix B. We remark that the definition of u simply means that we transport the vector (x, y) back to the vector $d\gamma^{-1}|_x y \in T_{\gamma^{-1}(x)}\mathcal{M}$ via the induced chart on $T\mathcal{M}$ and then we apply the exponential mapping, yielding a point $\exp_{\mathcal{M}}(\gamma^{-1}(x), (d\gamma)^{-1}|_x y) \in \mathcal{M}$. This point is then mapped to $\mathbb{R}^d$ via the chart γ which allows us to perform all operations with respect to a chart. The function v can be interpreted in an analogous way.

Our main result is that in any chart γ, we have the following approximation rate.

THEOREM 3.1 Assume that $f \in C^\alpha$ with $\alpha < m$ and $\alpha < s$. Then for $l < \alpha$ and any chart γ, we have

$$\left\| \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^l (\gamma \circ f - \gamma \circ Q^h f) \right\|_\infty = \mathcal{O}(h^{\alpha - l}). \tag{3.2}$$

In particular, the previous theorem easily implies that for *any* Riemannian metric g_l on $T^l\mathcal{M}$ with induced distance $\mathfrak{d}_l$ (meaning that $\mathfrak{d}_l$ measures the *geodesic distance* between two points; see DoCarmo, 1992), we have the following result.

THEOREM 3.2 Assume that $f \in C^\alpha$ with $\alpha < m$ and $\alpha < s$. Then for $l < \alpha$ and any Riemannian metric g_l on the vector bundle $T^l\mathcal{M}$ with induced distance metric $\mathfrak{d}_l$, we have the estimate

$$\sup_x \mathfrak{d}_l(d^l f, d^l Q^h f) = \mathcal{O}(h^{\alpha-l}), \tag{3.3}$$

where $d^l f : \mathbb{R} \to T^l\mathcal{M}$ denotes the total differential of order l (DoCarmo, 1992). The implicit constant is uniform for data values in a compact set.

Proof. Any chart γ induces the chart $d^l\gamma : T^l\mathcal{M} \to \mathbb{R}^{2^l d}$. With respect to this latter chart, by Theorem 3.1, we have

$$\|d^l\gamma \circ (d^l Q^h f(\cdot)) - d^l\gamma \circ (d^l f(\cdot))\|_\infty = \mathcal{O}(h^{\alpha-l}).$$

Since the inverse $(d^l\gamma)^{-1}$ of the chart $d^l\gamma$ is smooth and in particular Lipschitz with respect to the metric $\mathfrak{d}_l$, we arrive at the estimate

$$\begin{aligned}\sup_x \mathfrak{d}_l(d^l f(x), d^l Q^h f(x)) &= \sup_x \mathfrak{d}_l((d^l\gamma)^{-1} \circ d^l\gamma \circ (d^l f(x)), (d^l\gamma)^{-1} \circ d^l\gamma \circ (d^l Q^h f(x))) \\ &= \mathcal{O}(\|d^l\gamma \circ (d^l Q^h f(\cdot)) - d^l\gamma \circ (d^l f(\cdot))\|_\infty) = \mathcal{O}(h^{\alpha-l}),\end{aligned}$$

which proves the assertion for data contained in the domain of definition for any single chart γ. If the data values are contained in a compact set, finitely many charts cover this set and therefore the uniformity of the implied constant follows. □

3.2 *Sasaki metric*

Theorem 3.2 states that for *any* Riemannian metric defined on the tangent bundle $T^l\mathcal{M}$, we can show an approximation theorem as strong as for the linear case. It is a well-known fact that any differentiable manifold admits a Riemannian metric (Spivak, 1979). However, it would be nice to be able to single out one specific Riemannian metric on $T^l\mathcal{M}$ within which we can measure the approximation error between $d^l f$ and its approximation $d^l Q^h f$. The present section describes such a canonical metric, the so-called Sasaki metric which was introduced by Sasaki (1958). The reader who is content with the chart-dependent result in Section 3.1 may skip this section.

3.2.1 *Construction.* Before we can describe the Sasaki metric, we need some preliminary facts from the Riemannian geometry of tangent bundles; see Yano & Ishihara (1973) for more information. Assume that we are given a Riemannian manifold $(\mathcal{M}, g)$, where g is a symmetric $(0,2)$-tensor field on $\mathcal{M}$. Consider its tangent bundle $T\mathcal{M}$ which carries a natural manifold structure (DoCarmo, 1992). For $p \in \mathcal{M}$, we denote by $T_p\mathcal{M}$ the tangent space attached to p and likewise we denote by $T_{(p,u)}T\mathcal{M}$ the tangent space attached to the tangent bundle of $\mathcal{M}$ at the tangent vector $(p,u) \in T\mathcal{M}$. We want to find a suitable metric tensor g_T on the manifold $T\mathcal{M}$, meaning that g_T acts on pairs of tangent vectors of $T\mathcal{M}$.

With $\pi : T\mathcal{M} \to \mathcal{M}$ denoting the bundle projection, we define the *vertical subspace*

$$\mathcal{V}_{(p,u)} := \ker(d\pi|_{(p,u)})$$

of $T_{(p,u)}T\mathcal{M}$ at $(p,u) \in T\mathcal{M}$. The terminology 'vertical' stems from the fact that a vertical vector leaves the base point p stationary and only moves along the fibre $\pi^{-1}\{p\} \subset T\mathcal{M}$. Without going into detail, we would like to mention that the vertical subspace $\mathcal{V}_{(p,u)}$ can be complemented by the so-called *horizontal subspace*

$$\mathcal{H}_{(p,u)} := \ker(K_{(p,u)}),$$

where $K_{(p,u)} : T_{(p,u)}T\mathcal{M} \to T_p\mathcal{M}$ is defined as $(c(t), U(t)) \in T\mathcal{M}$, $(c(0), U(0)) = (p,u) \mapsto \nabla_{c'}U(0)$, where ∇ denotes the Levi-Civita connection of $(\mathcal{M}, g)$ (here, $\mathcal{H}_{(p,u)}$ consists of tangent vectors of parallel vector fields). We have

$$T_{(p,u)}T\mathcal{M} = \mathcal{H}_{(p,u)} \oplus \mathcal{V}_{(p,u)}$$

(Dombrowski, 1962; DoCarmo, 1992).

Definition 3.3 Let $X \in T_p\mathcal{M}$. Then the *horizontal lift* of X at $(p,u) \in T\mathcal{M}$ is the unique vector $X^h \in \mathcal{H}_{(p,u)}$ such that $d\pi_{(p,u)}(X^h) = X$. The *vertical lift* X^v of X at (p,u) is the unique vector $X^v \in \mathcal{V}_{(p,u)}$ such that $K_{(p,u)}(X^v) = X$.

Each $Z \in T_{(p,u)}T\mathcal{M}$ can be uniquely expressed as $Z = X^h + Y^v$ for $X, Y \in T_p\mathcal{M}$. Based on this decomposition, we can now define natural metrics on $T\mathcal{M}$ (Kappos, 2001).

Definition 3.4 A metric g_T on $T\mathcal{M}$ is called *natural* if

$$g_T(X^h, Y^h) = g(X,Y) \quad \text{and} \quad g_T(X^h, Y^v) = 0 \quad \text{for all vector fields } X, Y \text{ on } \mathcal{M}.$$

Several different natural metrics for $T\mathcal{M}$ can be defined by specifying conditions on $g_T(X^v, Y^v)$. For instance, the so-called *Cheeger–Gromoll metric* is uniquely defined by setting

$$g_T^{\mathrm{CG}}(X^v, Y^v)_{(p,u)} = \frac{1}{1 + g(X,Y)^2}(g(X,Y) + g(X,u)g(Y,u));$$

see Cheeger & Gromoll (1972). We will focus on the simpler *Sasaki metric*.

Definition 3.5 The Sasaki metric is the unique natural metric on $T\mathcal{M}$ which satisfies

$$g_T^{\mathrm{S}}(X^v, Y^v) = g(X,Y) \quad \text{for all vector fields } X, Y \text{ on } \mathcal{M}.$$

By iterating this construction, a Sasaki metric can be defined on the iterated tangent bundles $T^l\mathcal{M}$, $l \in \mathbb{Z}_+$.

The *geodesic distance metric* on $T^l\mathcal{M}$ induced by the Riemannian Sasaki metric will be denoted by $\mathfrak{s}_l$; it gives a natural notion of (local) distance $\mathfrak{s}_l(x,y)$ between points $x, y \in T^l\mathcal{M}$.

For a list of various properties satisfied by the Sasaki metric, we refer to the original paper Sasaki (1958).

3.2.2 *Intrinsic approximation results.* Now that we have singled out a canonical metric on the tangent bundles $T^l\mathcal{M}$, we can define smoothness properties for $\mathcal{M}$-valued functions intrinsically as follows.

DEFINITION 3.6 An $\mathcal{M}$-valued function $f : \mathbb{R} \to \mathcal{M}$ is in $C^\alpha(\mathbb{R}, \mathcal{M})$ if

$$\mathfrak{s}_{\lfloor\alpha\rfloor}(d^{\lfloor\alpha\rfloor}f(\cdot + h), d^{\lfloor\alpha\rfloor}f(\cdot)) = \mathcal{O}(h^{\alpha-\lfloor\alpha\rfloor})$$

with the implicit constant uniform in x for data values $f(x)$ in a compact set.

REMARK 3.7 It is easy to see that $f \in C^\alpha(\mathbb{R}, \mathcal{M})$ if and only if $\gamma \circ f \in C^\alpha$ for all charts γ.

We can finally state our main theorem.

THEOREM 3.8 Assume that $f \in C^\alpha(\mathbb{R}, \mathcal{M})$ with $\alpha < m$ and $\alpha < s$. Then for $l < \alpha$, we have the estimate

$$\mathfrak{s}_l(d^l f, d^l Q^h f) = \mathcal{O}(h^{\alpha-l}). \tag{3.4}$$

The implicit constant is uniform for data values $f(x)$ in a compact set.

Proof. The proof is a direct consequence of Theorem 3.2. □

REMARK 3.9 Although in Theorem 3.8 the implicit constant is only locally uniform, one can actually achieve a global constant whenever $\mathcal{M}$ is a homogeneous space, which is the case for all examples we mentioned in Section 1. Indeed, in that case one can always transform any part of $f(x)$ to a fixed domain of definition of a fixed chart and the constants become uniform.

REMARK 3.10 For computational reasons, it might be beneficial to replace the exponential function by a simpler retraction; see, for instance, Absil *et al.* (2008). It is easy to see that Theorem 3.8 also holds true in this case.

4. Proof of Theorem 3.1

From now on, we will simply write $Q^h f(\cdot)$ in place of $\gamma \circ Q^h f(\cdot)$ with the understanding that we are actually computing in a chart. The only property we shall use is the representation (3.1) and the smoothness of the functions u and v.

In a chart, the linear quasi-interpolation operator can be written as

$$\bar{Q}^h f(x) = \sum_{j\in\mathbb{Z}} \Phi(h^{-1}x - j)u(Q^h f(x), v(Q^h f(x), f(hj))). \tag{4.1}$$

The main idea is to compare the nonlinear approximation operator $Q^h f$ with its linear counterpart $\bar{Q}^h f$ and to use known properties of the linear operator to show the desired result. Estimating the error between linear and nonlinear quasi-interpolation will make up the bulk of this section, culminating in the proof of Theorem 3.1 in Section 4.2.

REMARK 4.1 The method of bounding the difference between a linear and a nonlinear approximation procedure has become a powerful tool in the study of nonlinear approximation schemes. It is often called the *method of proximity* and has mostly been used before in the analysis of nonlinear subdivision schemes; see, for instance Daubechies *et al.* (2004), Wallner & Dyn (2005), Wallner & Pottmann (2006), Xie & Yu (2008, 2010), Grohs (2009, 2010a,b), Grohs & Wallner (2009), Weinmann (2010).

We now define an important asymptotic quantity which will frequently be used in the sequel.

DEFINITION 4.2 For $f : \mathbb{R} \to \mathbb{R}^d$, $h > 0$, $l \in \mathbb{N}$, we define the quantity

$$\Omega_{l,h}(f) := \sum_{|\mathbf{r}|_1 = l} \prod_p \|\Delta_h^{r_p} f\|_\infty. \tag{4.2}$$

We collect some simple properties of the quantities $\Omega_{l,h}(f)$.

LEMMA 4.3 Assume that $f \in C^\alpha$ with $\alpha > l$. Then

$$\Omega_{l,h}(f) = \mathcal{O}(h^l). \tag{4.3}$$

Proof. Assume that $f \in C^\alpha$. Then, for any $l' < \alpha$, we have

$$\|\Delta_h^{l'} f\|_\infty = \mathcal{O}(h^{l'}).$$

It follows that

$$\Omega_{l,h}(f) = \sum_{|\mathbf{r}|_1 = l} \prod_p \|\Delta_h^{r_p} f\|_\infty = \mathcal{O}\left(\sum_{|\mathbf{r}|_1 = l} \prod_p h^{r_p}\right) = \mathcal{O}(h^l).$$

□

LEMMA 4.4 We have

$$\Omega_{l_1,h}(f)\Omega_{l_2,h}(f) = \mathcal{O}(\Omega_{l_1+l_2,h}(f)). \tag{4.4}$$

Proof. The proof follows simply by noting that all terms occurring on the left-hand side of (4.4) also occur on the right-hand side. □

4.1 *Proximity inequality*

Our basic strategy will be to compare the linear representation with the nonlinear one in order to infer properties for the nonlinear operator from properties of the linear one. Taylor expanding (3.1) up to order M, we obtain

$$\begin{aligned} Q^h f(x) = \sum_{k=0}^{M} \frac{1}{k!} d_2^{(k)} u|_{(Q^h f(x), 0}\left(\left[\sum_{j\in\mathbb{Z}} \Phi(h^{-1}x - j) v(Q^h f(x), f(hj))\right]^k\right) \\ + \frac{1}{(M+1)!} d_2^{(M+1)} u|_{\left(Q^h f(x), \xi\left(\sum_{j\in\mathbb{Z}} \Phi(h^{-1}x-j) v(Q^h f(x), f(hj))\right)\right)} \\ \times \left(\left[\sum_{j\in\mathbb{Z}} \Phi(h^{-1}x - j) v(Q^h f(x), f(hj))\right]^{M+1}\right). \end{aligned} \tag{4.5}$$

Taylor expanding (4.1) up to order M, we obtain

$$\bar{Q}^h f(x) = \sum_{k=0}^{M} \sum_{j\in\mathbb{Z}} \Phi(h^{-1}x - j) \frac{1}{k!} d_2^{(k)} u|_{(Q^h f(x),0)}([v(Q^h f(x), f(hj))]^k)$$
$$+ \sum_{j\in\mathbb{Z}} \Phi(h^{-1}x - j) \frac{1}{(M+1)!} d_2^{(M+1)} u|_{(Q^h f(x),\xi(v(Q^h f(x),f(hj))))}$$
$$\times ([v(Q^h f(x), f(hj))]^{M+1}). \tag{4.6}$$

We have used Taylor expansion in the second component, $d_2^{(k)}$ denoting the kth total derivative in the second coordinate and ξ a smooth function.

In what follows, we will make use of the following definition.

DEFINITION 4.5 For an arbitrary vector-valued sequence $p : \mathbb{Z} \to \mathbb{R}^d$ we define the lth forward difference via

$$\delta^l p(j) := \sum_{i=0}^{l} (-1)^{l-i} \binom{l}{i} p(j+i). \tag{4.7}$$

In addition, we shall use the following notation. For a vector $v \in \mathbb{R}^d$, we define the *concatenation*

$$[v]^k := (v, \dots, v) \in \mathbb{R}^{kd}.$$

Furthermore, for a function $f(x)$ and a multi-index $\mathbf{l} \in \mathbb{Z}_+^k$, we denote

$$\boldsymbol{\Delta}_h^{\mathbf{l}}[f]^k(\mathbf{x}) := (\Delta_h^{l_1} f(x_1), \dots, \Delta_h^{l_k} f(x_k)), \quad \mathbf{x} \in \mathbb{R}^k,$$

and similarly for a sequence $p(j)$,

$$\boldsymbol{\delta}^{\mathbf{l}}[p]^k(\mathbf{j}) := (\delta^{l_1} p(j_1), \dots, \delta^{l_k} p(j_k)), \quad \mathbf{j} \in \mathbb{Z}^k.$$

Our first goal is to gather useful estimates for the differences

$$F_k := d_2^{(k)} u|_{(Q^h f(x),0)} \left(\left[\sum_{j\in\mathbb{Z}} \Phi(h^{-1}x - j) v(Q^h f(x), f(hj)) \right]^k \right)$$
$$- \sum_{j\in\mathbb{Z}} \Phi(h^{-1}x - j) d_2^{(k)} u|_{(Q^h f(x),0)}([v(Q^h f(x), f(hj))]^k), \tag{4.8}$$

namely, the following lemma. The proof is based on a combinatorial argument exploiting the polynomial reproduction property (2.3) of Φ. Similar arguments have been used in previous work in different contexts (Daubechies *et al.*, 2004; Grohs, 2009; Xie & Yu, 2010).

LEMMA 4.6 For $l < s$, we have

$$|\Delta_{h'}^l F_k| = \mathcal{O}\left(\Omega_{m,h}(f) \sum_{i=0}^{l} (h^{-1}h')^{l-i} \Omega_{i,h'}(Q^h f) \right). \tag{4.9}$$

Proof. Define $\bar{j} := \lfloor h^{-1}x - N \rfloor$. Then every index j occurring in the summation formula for F_k satisfies $j \geqslant \bar{j}$ and $j - \bar{j} = \mathcal{O}(1)$, due to the support properties of Φ. By the definition of the forward differences, we can represent each vector $v(Q^h f(x), f(hj))$ by

$$v(Q^h f(x), f(hj)) = \sum_{i=0}^{j-\bar{j}} \binom{j-\bar{j}}{i} \delta^i v(Q^h f(x), f(h\bar{j})). \tag{4.10}$$

Inserting (4.10) into the definition of F_k, we obtain $F_k = F_k^1 + F_k^2$, where

$$F_k^1 := \sum_{i_1,\dots,i_k \in \mathbb{Z}_+} \sum_{j_1,\dots,j_k \in \mathbb{Z}} \prod_r \Phi(h^{-1}x - j_r) \binom{j_r - \bar{j}}{i_r}$$
$$\times\, d_2^{(k)} u|_{(Q^h f(x),0)} (\delta^{i_1} v(Q^h f(x), f(h\bar{j})), \dots, \delta^{i_k} v(Q^h f(x), f(h\bar{j})))$$

and

$$F_k^2 := \sum_{i_1,\dots,i_k \in \mathbb{Z}_+} \sum_{j \in \mathbb{Z}} \Phi(h^{-1} \cdot - j) \prod_r \binom{j-\bar{j}}{i_r}$$
$$\times\, d_2^{(k)} u|_{(Q^h f(x),0)} (\delta^{i_1} v(Q^h f(x), f(h\bar{j})), \dots, \delta^{i_k} v(Q^h f(x), f(h\bar{j}))).$$

Let us now fix $(i_1, \dots, i_k)$ with $i_1 + \dots + i_k < m$. Then the functions $p_r := p_r(j_r) = \binom{j_r - \bar{j}}{i_r}$, $p(j) := \prod_{r=1}^k \binom{j-\bar{j}}{i_r}$ are polynomials of total degree $< m$. Therefore, by polynomial reproduction of the function Φ of order $m-1$, we obtain

$$\sum_{j_1,\dots,j_k \in \mathbb{Z}} \prod_r \Phi(h^{-1}x - j_r) \binom{j_r - \bar{j}}{i_1} = \sum_{j \in \mathbb{Z}} \Phi(h^{-1}x - j) \prod_r \binom{j-\bar{j}}{i_r} = p(h^{-1}x),$$

and therefore we only need to consider terms with $i_1 + \dots + i_k \geqslant m$ in the summation formula for $F_k = F_k^1 - F_2^k$. In summary, F_k can be expressed as a finite linear combination of terms of the form

$$c(x) d_2^{(k)} u|_{(Q^h f(x),0)} (\delta^{i_1} v(Q^h f(x), f(h\bar{j})), \dots, \delta^{i_k} v(Q^h f(x), f(h\bar{j}))),$$

with

$$i_1 + \dots + i_k \geqslant m$$

and

$$c(x) := \prod_{r=1}^k \Phi(h^{-1}x - j_1) p_r(j_r) - \chi_{j_1 = \dots = j_k}(j_1, \dots, j_k) \Phi(h^{-1}x - j_1) p(j_1). \tag{4.11}$$

We need to bound $\Delta_{h'}^l F_k$ and by the above discussion it suffices to obtain a bound on

$$\Delta_{h'}^l c(x) d_2^{(k)} u|_{(Q^h f(x),0)} (\delta^{i_1} v(Q^h f(x), f(h\bar{j})), \dots, \delta^{i_k} v(Q^h f(x), f(h\bar{j}))) \tag{4.12}$$

with $i_1 + \cdots + i_k \geqslant m$. The expression (4.12) can be rewritten as a finite linear combination of terms of the form

$$(\Delta_{h'}^{l_1} c(x_1))(\Delta_{h'}^{l_2} d_2^{(k)} u|_{(Q^h f(x_2),0)})(\Delta_{h'}^{l_3} \delta^{i_1} v(Q^h f(x_3), f(h\bar{j})), \ldots, \Delta_{h'}^{l_{k+2}} \delta^{i_k} v(Q^h f(x_{k+2}), f(h\bar{j}))),$$

with $l_1 + \cdots + l_{k+2} = l$. By (4.11) and the smoothness of Φ we can estimate

$$\Delta_{h'}^{l_1} c(x) = \mathcal{O}(h^{-l_1}(h')^{l_1}). \tag{4.13}$$

Further, due to the smoothness of u and Lemma A2, we can estimate

$$\| \Delta_{h'}^{l_2} d_2^{(k)} u|_{(Q^h f(x_2),0)} \| = \mathcal{O}(\Omega_{l_2,h'}(Q^h f)). \tag{4.14}$$

Finally, by Lemma A3, we have for general l', i', that

$$\Delta_{h'}^{l'} \delta^{i'} v(Q^h f(x), f(h\bar{j})) = \mathcal{O}(\Omega_{l',h'}(Q^h f)\Omega_{i',h}(f)). \tag{4.15}$$

By putting estimates (4.13)–(4.15) into (4.12), we finally arrive at the desired result. □

We next treat the remainder terms in the Taylor representations.

LEMMA 4.7 For $M > 0$ and $l < s$, we have

$$\Delta_{h'}^{l} d_2^{(M+1)} u|_{(Q^h f(x), \xi(\sum_{j\in\mathbb{Z}} \Phi(h^{-1}x-j) v(Q^h f(x), f(hj))))} \left(\left[\sum_{j\in\mathbb{Z}} \Phi(h^{-1}x - j) v(Q^h f(x), f(hj)) \right]^{M+1} \right)$$
$$= \| \Delta_h f \|_\infty^{M-l} \mathcal{O}\left(\sum_{i=0}^{l} (h^{-1}h')^{l-i} \Omega_{i,h'}(Q^h f) \right) \tag{4.16}$$

and

$$\Delta_{h'}^{l} \sum_{j\in\mathbb{Z}} \Phi(h^{-1}x - j) \frac{1}{(M+1)!} d_2^{(M+1)} u|_{(Q^h f(x), \xi(v(Q^h f(x), f(hj))))} ([v(Q^h f(x), f(hj))]^{M+1})$$
$$= \| \Delta_h f \|_\infty^{M-l} \mathcal{O}\left(\sum_{i=0}^{l} (h^{-1}h')^{l-i} \Omega_{i,h'}(Q^h f) \right). \tag{4.17}$$

Proof. We start with (4.16). The proof goes by iteratively rewriting the divided differences in order to arrive at simpler expressions. First, note that the left-hand side of (4.16) can be expressed as a linear combination of terms of the form

$$(\Delta_{h'}^{l_1} d_2^{(M+1)} u|_{(Q^h f(x_1), \xi(\sum_{j\in\mathbb{Z}} \Phi(h^{-1}x_1 - j) v(Q^h f(x_1), f(hj))))})$$
$$\times \left(\boldsymbol{\Delta}_{h'}^{\mathbf{l_2}} \left[\sum_{j\in\mathbb{Z}} \Phi(h^{-1} \cdot -j) v(Q^h f(\cdot), f(hj)) \right]^{M+1} (\mathbf{x}_2) \right),$$

with $l_1 + |\mathbf{l_2}|_1 = l$. Due to the smoothness of u and (A.3), we can estimate

$$\|(\Delta_{h'}^{l_1} d_2^{(M+1)} u|_{(Q^h f(x), \xi(\sum_{j\in\mathbb{Z}} \Phi(h^{-1}x-j)v(Q^h f(x), f(hj))))})\|$$
$$\leqslant \sum_{r_1+r_2=l_1} \Omega_{r_1,h'}(Q^h f)\Omega_{r_2,h'}\left(\xi\left(\sum_{j\in\mathbb{Z}} \Phi(h^{-1}\cdot -j)v(Q^h f(\cdot), f(hj))\right)\right).$$

Due to the smoothness of ξ and (A.3), this expression can be further estimated by

$$\sum_{r_1+r_2=l_1} \Omega_{r_1,h'}(Q^h f) \sup_j \Omega_{r_2,h'}(\Phi(h^{-1}\cdot -j)v(Q^h f(\cdot), f(hj))). \tag{4.18}$$

For the second term in this product, we employ the following estimate for $r \geqslant 0$:

$$\|\Delta_{h'}^r \Phi(h^{-1}\cdot -j)v(Q^h f(\cdot), f(hj))\|_\infty \lesssim \sum_{t_1+t_2=r} \Omega_{t_1,h'}(\Phi(h^{-1}\cdot -j))\Omega_{t_2,h'}(v(Q^h f(\cdot), f(hj)))$$

which, by the smoothness of v and (A.3), can be bounded by

$$\sum_{t_1+t_2=r} \Omega_{t_1,h'}(\Phi(h^{-1}\cdot -j))\Omega_{t_2,h'}(Q^h f(\cdot)). \tag{4.19}$$

Due to the smoothness of Φ, this expression can be estimated by

$$\sum_{t_1+t_2=r} h^{-t_1}(h')^{t_1}\Omega_{t_2,h'}(Q^h f(\cdot)). \tag{4.20}$$

Combining (4.18) and (4.20) and using Lemma 4.4, we obtain

$$\|(\Delta_{h'}^{l_1} d_2^{(M+1)} u|_{(Q^h f(x), \xi(\sum_{j\in\mathbb{Z}} \Phi(h^{-1}x-j)v(Q^h f(x), f(hj))))})\| = \mathcal{O}\left(\sum_{i\leqslant l_1}(h^{-1}h')^{l_1-i}\Omega_{i,h'}(Q^h f)\right). \tag{4.21}$$

Now we go on to estimate entries in the vector

$$\boldsymbol{\Delta}_{h'}^{\mathbf{l_2}}\left[\sum_{j\in\mathbb{Z}} \Phi(h^{-1}\cdot -j)v(Q^h f(x), f(hj))\right]^{M+1};$$

since $|\mathbf{l_2}|_1 \leqslant l$, we can assume that only the first l entries of $\mathbf{l_2}$ are nonzero. The other $M+1-l$ entries can be bounded by $\mathcal{O}(\|\Delta_h f\|_\infty)$ by Lemma A1. For the other entries, we can use the estimate (4.20) with r replaced by $(l_2)_i$. Combining this estimate with (4.21) finally yields the desired estimate to show (4.16). The proof for (4.17) works analogously. □

Putting together the two previous estimates, we arrive at the following general result.

COROLLARY 4.8 We have the proximity inequality

$$\|\Delta^l_{h'}(Q^hf(\cdot)-\bar{Q}^hf(\cdot))\|_\infty=\mathcal{O}\left(\Omega_{m,h}(f)\sum_{i=0}^{l}(h^{-1}h')^{l-i}\Omega_{i,h'}(Q^hf)\right). \tag{4.22}$$

Proof. If we disregard the residue terms in the Taylor expansions of Q^hf and $\bar{Q}^hf$, then the estimate follows directly from Lemma 4.6. The residual terms are handled by Lemma 4.7 and by choosing $M \geqslant m+1+l$. □

4.2 *Main theorem*

We are almost in a position to give a proof of our main Theorem 3.1. First, we need the following lemma which states that the quasi-interpolants Q^hf are uniformly smooth and independent of h.

LEMMA 4.9 Assume that $f \in C^\alpha$ with $\alpha < m$ and $\alpha < s$. Then, for $l \leqslant \alpha$, we have

$$\|\Delta^l_{h'}Q^hf\|_\infty=\mathcal{O}((h')^l), \tag{4.23}$$

the implicit constant being independent of h.

Proof. We perform induction on l, the case $l=0$ being trivial. Let us now assume that for all $l'<l$, we have an inequality of the form (4.23). We can estimate

$$\|\Delta^l_{h'}Q^hf\|_\infty \leqslant \|\Delta^l_{h'}(Q^hf-\bar{Q}^hf)\|_\infty+\|\Delta^l_{h'}\bar{Q}^hf\|_\infty$$

which, by Corollary 4.8, can be bounded by

$$C\left(\sum_{|\mathbf{s}|_1=m}\prod_q\|\Delta^{s_q}_hf\|_\infty\sum_{|\mathbf{r}|_1\leqslant l}(h^{-1}h')^{l-|\mathbf{r}|_1}\prod_p\|\Delta^{r_p}_{h'}Q^hf\|_\infty\right)+\|\Delta^l_{h'}\bar{Q}^hf\|_\infty.$$

By the smoothness of f, Lemma 4.3, and the induction hypothesis, we can bound the above quantity by

$$Ch^{\alpha-l}(h')^l+Ch^\alpha\|\Delta^l_{h'}Q^hf\|_\infty+\|\Delta^l_{h'}\bar{Q}^hf\|_\infty,$$

with another constant C.

Utilizing the fact that $\|\Delta^l_{h'}\bar{Q}^hf\|_\infty=\mathcal{O}((h')^l)$, we arrive at the estimate

$$\|\Delta^l_{h'}Q^hf\|_\infty \leqslant Ch^{\alpha-l}(h')^l+Ch^\alpha\|\Delta^l_{h'}Q^hf\|_\infty+\bar{C}(h')^l.$$

Now let h be small enough such that $Ch^\alpha \leqslant \frac{1}{2}$. Then we have

$$\tfrac{1}{2}\|\Delta^l_{h'}Q^hf\|_\infty \leqslant Ch^{\alpha-l}(h')^l+\bar{C}(h')^l,$$

and this shows the desired assertion. □

We can finally conclude the proof of Theorem 3.1.

Proof of Theorem 3.1. The idea is to use the linear theory which states that

$$\left\| \left(\frac{\mathrm{d}}{\mathrm{d}x} \right)^l (f - \bar{Q}^h f) \right\|_\infty = \mathcal{O}(h^{\alpha - l}),$$

together with a suitable estimate for the difference between the linear and the nonlinear approximation procedure, namely, we will show that

$$\left\| \left(\frac{\mathrm{d}}{\mathrm{d}x} \right)^l (Q^h f - \bar{Q}^h f) \right\|_\infty = \mathcal{O}(h^{\alpha - l}). \tag{4.24}$$

In order to arrive at (4.24), we observe that the expression to be estimated can be written as

$$\left\| \left(\frac{\mathrm{d}}{\mathrm{d}x} \right)^l (Q^h f - \bar{Q}^h f) \right\|_\infty = \lim_{h' \to 0} (h')^{-l} \| \Delta_{h'}^l (Q^h f - \bar{Q}^h f) \|_\infty. \tag{4.25}$$

By Corollary 4.8, the right-hand side in (4.25) can be estimated by a constant times

$$(h')^{-l} \Omega_{m,h}(f) \sum_{i=0}^{l} (h^{-1} h')^{l-i} \Omega_{i,h'}(Q^h f). \tag{4.26}$$

By Lemma 4.9, we can estimate

$$\| \Delta_{h'}^r Q^h f \|_\infty = \mathcal{O}((h')^r), \quad r = 0, \ldots, l,$$

and hence

$$\Omega_{i,h'}(Q^h f) = \mathcal{O}((h')^i) \tag{4.27}$$

with an implicit constant independent of h and h'. Furthermore, due to the fact that $f \in C^\alpha$ and Lemma 4.3, we have

$$\Omega_{m,h}(f) = \mathcal{O}(h^\alpha). \tag{4.28}$$

Putting (4.27) and (4.28) into (4.26), we arrive at the estimate

$$(h')^{-l} \| \Delta_{h'}^l (Q^h f - \bar{Q}^h f) \|_\infty = \mathcal{O}(h^{\alpha - l}),$$

with the implicit constant independent of h'. By (4.25), this implies that

$$\left\| \left(\frac{\mathrm{d}}{\mathrm{d}x} \right)^l (Q^h f - \bar{Q}^h f) \right\|_\infty = \mathcal{O}(h^{\alpha - l}).$$

This proves (4.24) and hence the theorem. □

5. Numerical experiments

In the present section, we conduct some simple numerical experiments which confirm our theoretical findings. We will confine ourselves to the case $\mathcal{M} = \mathrm{SO}(2)$, the manifold of orthogonal 2×2 matrices with a positive determinant. This is a compact Riemannian manifold and also a Lie group. Its tangent bundle $T\mathcal{M}$ is given by $\mathrm{SO}(2) \times \mathfrak{so}_2$, where $\mathfrak{so}_2$ is the Lie algebra of 2×2 skew-symmetric matrices.

The exponential function of $\mathcal{M}$ is defined by the matrix exponential

$$\exp_{\mathcal{M}}(p,q) := p\left(\sum_{i=0}^{\infty} \frac{q^i}{i!}\right), \quad (p,q) \in \mathrm{SO}(2) \times \mathfrak{so}_2,$$

its (local) inverse is given by

$$\log_{\mathcal{M}}(p,q) := \sum_{i=0} (-1)^i \frac{(qp^{-1} - I)^i}{i}, \quad p,q \in \mathrm{SO}(2),$$

the usual matrix logarithm.

We compute the Riemannian centre of mass

$$\mathrm{av}_{\mathcal{M}}((p(j))_j, (w(j))_j)$$

via the fixed-point iteration

$$x_{n+1} = \exp_{\mathcal{M}}\left(x_n, \sum_j w(j) \log_{\mathcal{M}}(x_n, p(j))\right).$$

In Karcher (1977), it is shown that this iterative procedure converges linearly to the centre of mass.

REMARK 5.1 In our simple experiments, we have used the simple fixed-point iteration defined above. For more numerically demanding applications, one could replace this iteration by a Newton-type scheme (Groisser, 2004). Also it is possible to replace the exponential function by another, possibly more efficiently computable, retraction; see, e.g., Absil *et al.* (2008).

EXAMPLE 5.2 (Quasi-interpolation with cubic B-splines: smooth data) In this example, we set

$$\Phi(x) := -\frac{1}{6}B_3(x-1) + \frac{4}{3}B_3(x) - \frac{1}{6}B_3(x+1),$$

with B_3 the cardinal cubic B-spline function (Chui, 1992). In the linear case, this gives a well-known quasi-interpolation scheme with polynomial reproduction $m = 4$ and smoothness $s = 2$. We study the approximation speed for the smooth SO(2)-valued function

$$f(x) = \begin{pmatrix} \cos(\sin(2x)) & -\sin(\sin(2x)) \\ \sin(\sin(2x)) & \cos(\sin(2x)) \end{pmatrix} \tag{5.1}$$

and its first two derivatives; see Fig. 1. The error is measured in the Frobenius norm. Our experiment confirms the approximation rates predicted by the theory.

EXAMPLE 5.3 (Quasi-interpolation with cubic B-splines: nonsmooth data) In this example, we study the same approximation procedure as in Example 5.2, this time with the nonsmooth function $f : [0,1] \to$

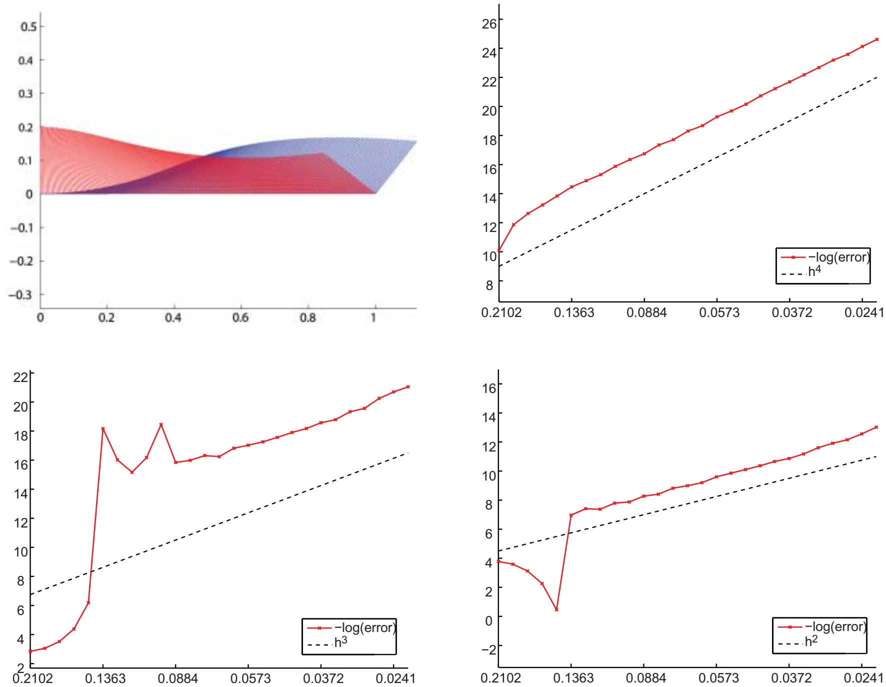

FIG. 1. In this example, the function $f(x):[0,1]\to \mathrm{SO}(2)$ given by (5.1) is approximated by nonlinear cubic B-spline quasi-interpolation. Top left: $f(x)$. A matrix in SO(2) is illustrated by its two orthogonal row vectors. Top right: approximation error, plotted against h with log–log axis scaling. The dashed line represents the ruling line corresponding to an approximation rate h^4. Bottom left: approximation error of first derivatives, plotted against h with log–log axis scaling. The dashed line represents the ruling line corresponding to an approximation rate h^3. Bottom right: approximation error of second derivatives, plotted against h with log–log axis scaling. The dashed line represents the ruling line corresponding to an approximation rate h^2.

SO(2) defined via

$$f(x) = \begin{pmatrix} \cos(|x-0.5|^{1/2}) & -\sin(|x-0.5|^{1/2}) \\ \sin(|x-0.5|^{1/2}) & \cos(|x-0.5|^{1/2}) \end{pmatrix}. \tag{5.2}$$

The function f is only in $C^{1/2}$ and therefore we only expect an approximation rate of $\frac{1}{2}$ which is observed in Fig. 2.

EXAMPLE 5.4 (High-order approximation) In our final example, we consider quasi-interpolation with a quintic B-spline function. With B_5 the cardinal B-spline function of degree 5, we put

$$\Phi(\cdot) = \frac{13}{240}B_5(x-2) - \frac{7}{15}B_5(x-1) + \frac{73}{40}B_5(x) - \frac{7}{15}B_5(x+1) + \frac{13}{240}B_5(x+2).$$

It can be shown that this function satisfies the assumptions (2.1)–(2.4) with $m=6$ and $s=4$. It follows that the approximation error

$$f - Q^h f$$

is expected to be of order h^6. This is confirmed by the numerical experiment in Fig. 3, where the smooth function f from Example 5.2 is approximated numerically.

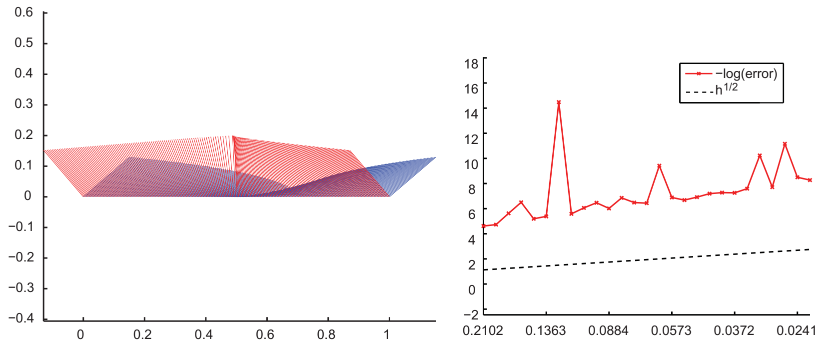

FIG. 2. In this example, the function $f(x):[0,1]\to \mathrm{SO}(2)$ given by (5.2) is approximated by nonlinear cubic B-spline quasi-interpolation. Left: $f(x)$. Right: approximation error, plotted against h with log–log axis scaling. The dashed line represents the ruling line corresponding to an approximation rate $h^{1/2}$.

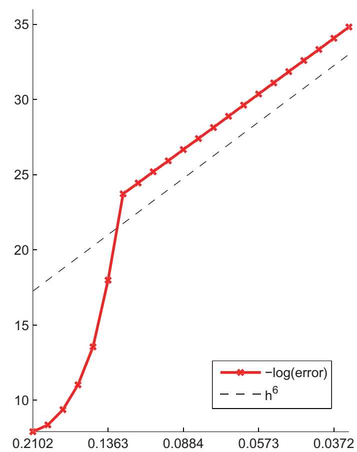

FIG. 3. Approximation rate of the smooth function f given in (5.1) by quintic B-spline quasi-interpolation. The approximation error is plotted against h with log–log axis scaling. The dashed line represents the ruling line corresponding to an approximation rate h^6; the plot suggests the expected approximation order h^6.

Acknowledgements

The author thanks Oliver Sander and Johannes Wallner for useful comments.

Funding

This research was supported by the European Research Council under grant ERC AdG 247277.

REFERENCES

ABSIL, P., MAHONY, R. & SEPULCHRE, R. (2008) *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton University Press.

BRAESS, D. (2007) *Finite Elements*. Cambridge, UK: Cambridge University Press.

CHEEGER, J. & GROMOLL, D. (1972) On the structure of complete manifolds of nonnegative curvature. *Ann. Math.*, **96**, 413–443.

CHUI, C. (1992) *An Introduction to Wavelets*. New York, NY: Academic Press.

DAUBECHIES, I., RUNBORG, O. & SWELDENS, W. (2004) Normal multiresolution approximation of curves. *Construct. Approx.*, **20**, 399–463.

DE BOOR, C. & FIX, G. (1973) Spline approximation by quasiinterpolants. *J. Approx. Theory*, **8**, 19–45.

LE BIHAN, D., MANGIN, J.-F., POUPON, C., CLARK, C., PAPPATA, S., MOLKO, N. & CHABRIAT, H. (2001) Diffusion tensor imaging: concepts and applications. *J. Magn. Resonance Imag.*, **13**, 534–546.

DEVORE, R. & LORENTZ, G. (1993) *Constructive Approximation*. Berlin: Springer.

DOCARMO, M. (1992) *Riemannian Geometry*. Basel: Birkhäuser.

DOMBROWSKI, P. (1962) On the geometry of the tangent bundle. *J. Reine Angewandte Math.*, **210**, 73–88.

GROHS, P. (2009) Smoothness of interpolatory multivariate subdivision in Lie groups. *IMA J. Numer. Anal.*, **29**, 760–772.

GROHS, P. (2010a) Approximation order from stability of nonlinear subdivision schemes. *J. Approx. Theory*, **162**, 1085–1094.

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

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

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

GROISSER, D. (2004) Newton's method, zeroes of vector fields, and the Riemannian center of mass. *Adv. Appl. Math.*, **33**, 95–135.

HARIZANOV, S. & OSWALD, P. (2010) Stability of nonlinear subdivision and multiscale transforms. *Construct. Approx.*, **31**, 359–393.

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

KAPPOS, E. (2001) Natural metrics on tangent bundles. *Master's Thesis*, Lund University, Sweden.

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

PENNEC, X., FILLARD, P. & AYACHE, N. (2006) A Riemannian framework for tensor computing. *Int. J. Comput. Vis.*, **66**, 41–66.

RAHMAN, I. U., DRORI, I., STODDEN, V. C., DONOHO, D., & SCHRÖDER, P. (2006) Multiscale representations for manifold-valued data. *Multiscale Model. Simul.*, **4**, 1201–1232.

SANDER, O. (2010) Geodesic finite elements for Cosserat rods. *Int. J. Numer. Methods Eng.*, **82**, 1645–1670.

SANDER, O. (2011) Geodesic finite elements on simplicial grids, doi: 10.1002/nme.4366.

SASAKI, S. (1958) On the differential geometry of tangent bundles of Riemannian manifolds. *Tohoku Math. J.*, **10**, 338–354.

SPIVAK, M. (1979) *A Comprehensive Introduction to Differential Geometry*, vol. 1. Publish or Perish. Michigan.

WALLNER, J. & DYN, N. (2005) Convergence and C^1 analysis of subdivision schemes on manifolds by proximity. *Comput. Aided Geom. Des.*, **22**, 593–622.

WALLNER, J., NAVA YAZDANI, E., & WEINMANN, A. (2011) Convergence and smoothness analysis of subdivision rules in Riemannian and symmetric spaces. *Adv. Comput. Math.*, **34**, 201–218.

WALLNER, J. & POTTMANN, H. (2006) Intrinsic subdivision with smooth limits for graphics and animation. *ACM Trans. Graph.*, **25**, 274–356.

WEINMANN, A. (2010) Nonlinear subdivision schemes on irregular meshes. *Construct. Approx.*, **31**, 395–415.

XIE, G. & YU, T. P.-Y. (2008) Smoothness equivalence properties of general manifold-valued data subdivision schemes. *Multiscale Model. Simul.*, **7**, 1073–1100.

XIE, G. & YU, T. (2010) Smoothness equivalence properties of interpolatory Lie group subdivision schemes. *IMA J. Numer. Anal.*, **30**, 731–750.

Xie, G. & Yu, T. P.-Y. (2011) Approximation order equivalence properties of manifold-valued data subdivision schemes. *IMA J. Numer. Anal*, 32, 687–700.

Yano, K. & Ishihara, S. (1973) *Tangent and Cotangent Bundles: Differential Geometry*, vol. 16. New York, NY: Marcel Dekker.

Appendix A. Additional technical results

Lemma A1 We have

$$\chi_{[-N,N]}(h^{-1}x-i)v(Q^h(x),f(hj))=\mathcal{O}(\|\Delta_h f(\cdot)\|_\infty).$$

Proof. Let us assume for simplicity that $\mathcal{M}$ is embedded in Euclidean space. First note that

$$\log_{\mathcal{M}}(x,y)=\log_{\mathcal{M}}(x,x)+\mathcal{O}(|x-y|)=\mathcal{O}(|x-y|) \tag{A.1}$$

since $\log(x,x)=0$. By Karcher (1977, 1.5.1), it holds that

$$|f(hj)-Q^hf(x)|\lesssim \mathfrak{d}(f(hj),Q^hf(x))\lesssim\left|\sum_{l\in\mathbb{Z}}\Phi(h^{-1}x-l)\log_{\mathcal{M}}(f(hl),f(hi))\right|. \tag{A.2}$$

This expression is only nonzero if

$$h^{-1}x-l\in[-N,N].$$

Further, by assumption, we have

$$h^{-1}x-i\in[-N,N],$$

which implies that $|hl-hi|\lesssim h$. Therefore, by (A.1) and (A.2), we obtain

$$\chi_{[-N,N]}(h^{-1}x-i)|f(hj)-Q^hf(x)|=\mathcal{O}(\|\Delta_h f\|_\infty).$$

The final estimate follows from noting that

$$|v(f(hj),Q^hf(x))|\lesssim|f(hj)-Q^hf(x)|,$$

which can be shown in the same fashion as (A.1). □

Lemma A2 Assume that $g:\mathbb{R}^k\to\mathbb{R}$ is a smooth function and $\mathbf{f}=(f_1,\ldots,f_k):\mathbb{R}\to\mathbb{R}^k$. Then

$$\Delta_{h'}^l g(\mathbf{f}(\cdot))=\mathcal{O}\left(\sum_{|\mathbf{r}|_1=l}\prod_q \Omega_{r_q,h'}(f_q)\right) \tag{A.3}$$

Proof. We use Taylor expansion of g at $(f_1(x), \ldots, f_k(x))$:

$$g(\mathbf{f}(x+y)) = \sum_{i=1}^{l-1} \sum_{\mathbf{s}\in\{1,\ldots,k\}^i} d_{\mathbf{s}}^i g|_{(\mathbf{f}(x))}(f_{s_1}(x+y) - f_{s_1}(x), \ldots, f_{s_i}(x+y) - f_{s_i}(x))$$
$$+ \sum_{\mathbf{s}\in\{1,\ldots,k\}^l} d_{\mathbf{s}}^l g|_{\xi(\mathbf{f}(x+y))}(f_{s_1}(x+y) - f_{s_1}(x), \ldots, f_{s_i}(x+y) - f_{s_i}(x)).$$

Here, the expression $d_{\mathbf{s}}^i g$ denotes the partial derivative $\partial_{s_1} \ldots \partial_{s_i} g$. The last term in the above expression is clearly of order

$$\mathcal{O}\left(\sum_{\mathbf{s}\in\{1,\ldots,k\}^l} \prod_q \|\Delta_{h'} f_{s_q}\|_\infty^l \right) \quad \text{for } |y| \leqslant lh',$$

and therefore we can estimate

$$\Delta_{h'}^l g(\mathbf{f}(x)) = \sum_{i=1}^{l-1} \sum_{\mathbf{s}\in\{1,\ldots,k\}^i} \Delta_{h'}^l d_{\mathbf{s}}^i g|_{\mathbf{f}}(f_{s_1}(x+y) - f_{s_1}(x), \ldots, f_{s_i}(x+y) - f_{s_i}(x))$$
$$+ \mathcal{O}\left(\sum_{\mathbf{s}\in\{1,\ldots,k\}^l} \prod_q \|\Delta_{h'} f_{s_q}\|_\infty^l \right). \tag{A.4}$$

Each of the terms

$$\Delta_{h'}^l d_{\mathbf{s}}^i g|_{\mathbf{f(x)}}(f_{s_1}(x+y) - f_{s_1}(x), \ldots, f_{s_i}(x+y) - f_{s_i}(x))$$

can be expressed as a finite linear combination of terms of the form

$$d_{\mathbf{s}}^i g|_{\mathbf{f}(x)}(\Delta_{h'}^{l_1} f_{s_1}(x_1 + \cdot) - f_{s_1}(x_1), \ldots, \Delta_{h'}^{l_i} f_{s_i}(x_i + \cdot) - f_{s_i}(x_i))$$

with $l_1 + \cdots + l_i = l$ and some $(x_1, \ldots, x_i) \in \mathbb{R}^i$. This gives the final bound. □

LEMMA A3 We have the estimate

$$\Delta_{h'}^{l_1} \delta^{l_2} v(Q^h(x), f(hj)) = \mathcal{O}(\Omega_{l_1,h'}(Q^h f)\Omega_{l_2,h}(f)). \tag{A.5}$$

Proof. Again we use Taylor expansion of the bivariate function $v(x,y)$ together with arguments akin to the proof of Lemma A2. Put $\bar{j} = \lfloor h^{-1}x - N \rfloor$. Then $h\bar{j} - hj = \mathcal{O}(h)$ for all j such that $v(Q^h f(x), f(hj)) \neq 0$. We have

$$v(Q^h f(x), f(hj)) = \sum_{i=0}^{l_1-1} d_2^{(i)} v|_{(Q^h f(x), f(h\bar{j}))}([f(hj) - f(h\bar{j})]^i)$$
$$+ d_2^{(l_2)} v|_{(Q^h f(x), \xi(f(hj)))}([f(hj) - f(h\bar{j})]^i).$$

For any i, we can write

$$\delta^{l_2} \Delta_{h'}^{l_1} d_2^{(i)} v|_{(Q^h f(x), f(h\bar{j}))}([f(hj) - f(h\bar{j})]^i)$$

as a finite linear combination of terms of the form

$$\Delta_{h'}^{l_1} d_2^{(i)} v|_{(Q^h f(x), f(h\bar{j}))}(\boldsymbol{\delta}^{\mathbf{s}}[f(hj) - f(h\bar{j})]^i)$$

with $|\mathbf{s}|_1 = l_2$. This expression can be bounded by

$$\|\Delta_{h'}^{l_1} d_2^{(i)} v|_{(Q^h f(x), f(h\bar{j}))}\| \sum_{|\mathbf{s}|_1 = l_2} \prod_q \|\Delta_h^{s_q} f\|_\infty.$$

Furthermore, by Lemma A2,we have

$$\|\Delta_{h'}^{l_1} d_2^{(i)} v|_{(Q^h f(x), f(h\bar{j}))}\| = \mathcal{O}\left(\sum_{|\mathbf{r}|_1 = l_1} \prod_p \|\Delta^{r_p} Q^h f\|_\infty\right).$$

To handle the residual term, we note that we can bound the expression

$$\delta^{l_2} \Delta_{h'}^{l_1} d_2^{(l_2)} v|_{(Q^h f(x), \xi(f(hj)))}([f(hj) - f(h\bar{j})]^i)$$

by terms of the form

$$\Delta_{h'}^{l_1} d_2^{(l_2)} v|_{(Q^h f(x), \xi(f(hj)))}([f(hj) - f(h\bar{j})]^i),$$

which can in turn be bounded by

$$\|\Delta_{h'}^{l_1} d_2^{(l_2)} v|_{(Q^h f(x), \xi(f(hj)))}\| |f(hj) - f(h\bar{j})|^{l_2} = \mathcal{O}\left(\sum_{|\mathbf{r}|_1 = l_1} \prod_p \|\Delta^{r_p} Q^h f\|_\infty \|\Delta_h f\|_\infty^{l_2}\right).$$

We have used Lemma A2 and the fact that $hj - h\bar{j} = \mathcal{O}(h)$. Summing these estimates gives the desired result. □

Appendix B. Some basic notions of Riemannian geometry

For the convenience of the reader, we summarize a few basic definitions in Riemannian geometry. Everything covered here is completely elementary and can be found in any textbook on differential geometry, such as the work of DoCarmo (1992). This appendix is meant to help the reader who is not familiar with Riemannian geometry with the understanding of the results of this paper, especially Section 3.

B.1 *Smooth manifolds and tangent space*

We start with a smooth manifold $\mathcal{M}$, which can be defined as a Hausdorff topological space which 'locally looks like Euclidean space' in the sense that there exists a family of charts $\gamma : U_\gamma \subset \mathcal{M} \to \gamma(U_\gamma) \subset \mathbb{R}^d$ such that U_γ is an open set and

- the functions $\gamma : U_\gamma \to \gamma(U_\gamma)$ are bijective,
- the composition of any two charts $\gamma_1 \circ \gamma_2^{-1}$ is smooth as a function defined on a subset of $\mathbb{R}^d$ mapping into $\mathbb{R}^d$,
- for each $p \in \mathcal{M}$ there exists a chart γ such that $p \in U_\gamma$.

A smooth manifold possesses enough structure to define the notion of a smooth curve; a curve $c(t)$: $[-1,1] \to \mathcal{M}$ is smooth if the mapping $\gamma \circ c(t) : [-1,1] \to \mathbb{R}^d$ is smooth for all charts whose domain of definition contains the image of c in $\mathcal{M}$. For an arbitrary point $p \in \mathcal{M}$, one can define the *tangent space* in p as consisting of all smooth curves $c : [-1,1] \to \mathcal{M}$ with $c(0) = p$ modulo the equivalence relation $c_1 \sim c_2 :\Leftrightarrow (\mathrm{d}/\mathrm{d}t)\gamma \circ c_1(0) = (\mathrm{d}/\mathrm{d}t)\gamma \circ c_2(0)$ for one (and then all) chart(s) γ with $p \in U_\gamma$. We shall denote the equivalence class of a curve c by $[c]$. This means that tangent vectors in p are by definition velocities of curves through p. These form a vector space $T_p\mathcal{M}$, the tangent space of $\mathcal{M}$ in p. The *full tangent bundle* is simply the union of all spaces $T_p\mathcal{M}$ over all points p: $T\mathcal{M} := \bigcup_{p\in\mathcal{M}} T_p\mathcal{M}$. Sometimes, we use the somewhat imprecise notation (p, u) for a tangent vector attached to $p \in \mathcal{M}$, in order to mark its base point p.

Now if we have two manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$ and a function $f : \mathcal{M}_1 \to \mathcal{M}_2$, then we can define the *derivative* of f as follows:

$$\mathrm{d}f : \begin{cases} T\mathcal{M}_1 \to T\mathcal{M}_2, \\ [c] \in T_p\mathcal{M}_1 \mapsto [f \circ c] \in T_{f(p)}\mathcal{M}_2. \end{cases}$$

Restricted to the tangent space $T_p\mathcal{M}_1$ at a point $p \in \mathcal{M}_1$, the mapping $\mathrm{d}f$ is linear. We denote this restriction by $\mathrm{d}f|_p : T_p\mathcal{M}_1 \to T_{f(p)}\mathcal{M}_2$. Using these notions, it is possible to equip the tangent space $T\mathcal{M}$ of a smooth manifold $\mathcal{M}$ with the structure of a smooth manifold: given any chart $\gamma : U_\gamma \to \mathbb{R}^d$, one can define the *induced chart* $(\gamma, \mathrm{d}\gamma) : TU_\gamma \to T\mathbb{R}^d \cong \mathbb{R}^{2d}$ by mapping a tangent vector $[c] \in T_p\mathcal{M}$ to the tangent vector $(\mathrm{d}/\mathrm{d}t)(\gamma \circ c)(0)$ in $\mathbb{R}^d$, attached to $\gamma(p) \in \mathbb{R}^d$. It can be shown that the system of induced charts equips $T\mathcal{M}$ with the structure of a smooth manifold. Therefore, one can also speak of *iterated tangent spaces* $T^l\mathcal{M}$ which carry the structure of a smooth manifold. The tangent space $T\mathcal{M}$ carries the structure of a *vector bundle*, which means that locally it looks like a parametrized family of vector spaces attached to different points. We denote the *bundle projection* $\pi : T\mathcal{M} \to \mathcal{M}$ defined by $\pi([c]) := c(0)$. It projects a tangent vector onto its base point.

B.2 *Riemannian structure*

Intuitively, a *Riemannian manifold* is a smooth manifold where one can measure distances much in the same way as in Euclidean space. For instance, to measure the length of a curve $c : [a, b] \to \mathcal{M}$, one can define the length of c as the integral $\int_a^b \langle \mathrm{d}c(t), \mathrm{d}c(t)\rangle_{T_{c(t)}\mathcal{M}} \mathrm{d}t$. For this definition of arc length, it is necessary to measure lengths of velocity vectors. This motivates the definition of a Riemannian manifold as a smooth manifold $\mathcal{M}$, together with a smooth family $(\langle\cdot,\cdot\rangle_p)_{p\in\mathcal{M}}$ of Euclidean inner products $\langle\cdot,\cdot\rangle_p$ on the vector spaces $T_p\mathcal{M}$. Often one uses the notation $\langle\cdot,\cdot\rangle_p := g(\cdot,\cdot)_p$. The expression g can be interpreted as a family of 2-tensors, parametrized by $p \in \mathcal{M}$. Therefore, one calls g a (symmetric) $(0, 2)$-*tensor field* and refers to the pair $(\mathcal{M}, g)$ as a Riemannian manifold. In a Riemannian manifold, we are able to measure the length of a curve c connecting two points $p, q \in \mathcal{M}$ and this readily introduces a notion of *induced distance* $\mathfrak{d}(p, q)$ between two points $p, q \in \mathcal{M}$ as the infimal arc length over all curves connecting p and q.

A *geodesic* can be defined as being the curve with minimal arc length connecting two points. It can be shown that this definition is locally unique. Geodesics satisfy a second-order autonomous ODE and are as such fully determined by their initial value and velocity. Therefore, one can define the *exponential mapping* $\exp_\mathcal{M}$ which maps a tangent vector $[c]$, attached to $p \in \mathcal{M}$, to the point $\exp_\mathcal{M}(p)$ obtained from following the unique geodesic emanating from p with initial velocity $[c]$ for one time instant. As such, the exponential mapping is a mapping from $T\mathcal{M}$ to $\mathcal{M}$. It can be shown that locally there exists an inverse to $\exp_{\mathcal{M}}$ which maps two points $p, q \in \mathcal{M}$ to the initial velocity of the geodesic emanating from p and connecting p and q. This mapping is called the *logarithm mapping* and it is a mapping from (a subset of) $\mathcal{M} \times \mathcal{M}$ to $T\mathcal{M}$.

Another important concept in Riemannian geometry is the *Levi-Civita connection* which provides a geometric way to identify tangent spaces attached to different points. We refer to DoCarmo (1992) for details and simply mention that the Levi-Civita connection allows vector fields (i.e., mappings $V : \mathcal{M} \to T\mathcal{M}, \pi(V(p)) = p$) to be differentiated along curves $c : [a, b] \to \mathcal{M}$ yielding the expression $\nabla_{c'} V$, where c' denotes the differential of c.