Diss. ETH No. 23579

Numerical Methods for Optimization and Variational Problems with Manifold-Valued Data

A dissertation submitted to
ETH Zürich

for the degree of
Doctor of Sciences

presented by

MARKUS SPRECHER

MSc ETH Math, ETH Zürich born August 30, 1986
citizen of Chur, GR

accepted on the recommendation of
Prof. Dr. Philipp Grohs, ETH Zürich, examiner
Prof. Dr. Oliver Sander, TU Dresden, co-examiner
Prof. Dr. Johannes Wallner, TU Graz, co-examiner

2016

Acknowledgments

I am thankful to everyone who has contributed to the success of this thesis in any way:

Prof. Dr. Philipp Grohs for giving me the opportunity to do my PhD at ETH Zürich and work on a very interesting and new topic. Thanks to the scientific freedom and trust I experienced by working with him, it has been a great time of acquiring new knowledge among different areas of mathematics. I admire his mathematical intuition and also his broad knowledge.

Prof. Dr. Oliver Sander and Prof. Dr. Johannes Wallner for acting as co-examiners.

Zeljko Kereta and Andreas Bärtschi for proof-reading my thesis.

All the members of the Seminar for applied mathematics who make this institute not only a good place to work but also to live. I enjoyed the mathematical and non-mathematical activities we spend together.

My parents giving me the opportunity to spend many weekends in their beautiful and quiet environment in close touch with nature.

Dominik Zippert and Andreas Bärtschi for supporting me during all these years.

Manuel Cavegn, Mara Nägelin, Matthias Wellershoff and Pascal Debus for their fruitful collaboration as part of their bachelor or master thesis.

My brother Andi. The activities we spent together were always a welcome change.

My brother Daniel and his wife Déborah for setting good examples and supporting me during all these years.

A mathematical theory is not
to be considered complete until you
have made it so clear that you can
explain it to the first man whom
you meet on the street.

DAVID HILBERT (1862-1943)

Abstract

In this thesis, we consider optimization and variational problems where the data is constrained to lie on a Riemannian manifold. Two examples, we will particularly focus on, are the denoising of manifold-valued images by minimizing a total variation (TV) functional and the minimization of the harmonic energy with prescribed boundary data. Typical examples for the manifold in these applications are the sphere S^n (e.g. for the chromaticity part of an RGB-image, or unit vector fields), the special orthogonal group $SO(n)$ (e.g. for rigid body motion) or the set of positive definite matrices $SPD(n)$ (e.g. for diffusion tensor magnetic resonance imaging (DT-MRI)).

For the optimization problems, we will use techniques of Absil et al. [3], which generalize many optimization techniques for functionals on $\mathbb{R}^n$ to optimization techniques for functionals on manifolds. We present a theory which shows how these techniques can be applied to our problems.

To minimize the TV functional, we propose an iteratively reweighted minimization (IRM) algorithm, which is an adaptation of the well-known iteratively reweighted least squares (IRLS) algorithm. We show that the algorithm can be applied to Hadamard manifolds and the half-sphere.

To minimize the harmonic energy, we use a natural discretization. As it turns out, this discretization has the same structure as the functional occuring in the IRM algorithm. This will allow us to reuse derived results. In particular, it follows that the discretization of the harmonic energy has a unique minimizer. For the half-sphere we can prove convergence of the discrete minimizer towards the minimizer of the harmonic energy. We will also present a general technique to numerically solve variational problems with manifold valued data by minimizing the functional on a subspace. This subspace is constructed from a classical "finite element space". Minimizing a functional over the subspace will reduce to an optimization problem on a Cartesian power of the manifold. To estimate the discretization error, we will derive a nonlinear Céa lemma showing that the discretization error can be bounded by the best approximation error. To estimate the best approximation error, we generalize a class of approximation operators into finite element spaces and show that the generalization satisfies the same error estimate as its linear counterpart.

The thesis can be summarized by saying that we generalize numerical methods and theories for optimization and variational problems from the real-valued case to the manifold-valued case.

Zusammenfassung

In dieser Dissertation betrachten wir Optimierungs- und Variationsprobleme mit der Nebenbedingung, dass die Daten auf einer Riemannschen Mannigfaltigkeit liegen müssen. Zwei Beispiele, denen wir uns besonders widmen, sind das Entrauschen von mannigfaltigkeitswertigen Bildern mittels Minimierung eines Variationsfunktionals und die Minimierung der harmonischen Energie mit vorgebenen Randdaten. Typische Beispiele für die Mannigfaltigkeit in diesen Anwendungen sind die Sphäre S^n (z.B. für den Chromatizitätsanteil eines RGB-Bildes, oder Einheitsvektorfelder), die spezielle orthogonale Gruppe $SO(n)$ (z.B. für starre Bewegungen) oder die Menge der positiv definiten Matrizen $SPD(n)$ (z.B. für die Diffusions-Tensor-Bildgebung (DT-MRI)).

Für die Optimierungsprobleme verwenden wir Techniken von Absil et al. [3], welche viele Optimierungstechniken für Funktionale auf $\mathbb{R}^n$ zu Optimierungstechniken für Funktionale auf Mannigfaltigkeiten verallgemeinern. Wir leiten eine Theorie her um diese Techniken auf unsere Probleme anwenden zu können.

Um das Variationsfunktional zu minimieren, schlagen wir einen iterativen Neugewichtungsalgorithmus (IRM) vor, welcher eine Abwandlung des allgemein bekannten iterativen kleinste Quadrate Neugewichtungsalgorithmus (IRLS) ist. Wir zeigen, dass unser Algorithmus auf Hadamardmannigfaltigkeiten und die Halbkugel anwendbar ist.

Um die harmonische Energie zu minimieren, verwenden wir eine natürliche Diskretisierung. Wie sich zeigen wird hat diese Diskretisierung die gleiche Struktur wie das Funktional, welches beim IRM-Algorithmus auftritt. Dies wird uns erlauben, hergeleitete Resultate wiederzuverwenden. Unter anderem folgt dann, dass die Diskretisierung der harmonischen Energie einen eindeutigen Minimierer hat. Für die Halbkugel können wir zeigen, dass dieser Minimierer gegen den Minimierer der harmonischen Energie konvergiert. Wir werden auch eine allgemeine Methode präsentieren, um numerisch Variationsprobleme mit mannigfaltigkeitswertigen Daten zu lösen. Die Methode basiert auf dem Lösen des Minimierungsproblems auf einem Unterraum. Dieser Unterraum stammt jeweils von einem “Finite Elemente Raum” ab. Minimierung eines Funktionals über diesen Unterraum reduziert sich auf die Minimierung eines Funktionals auf dem kartesischen Produkt einer Mannigfaltigkeit. Um den Diskretisierungsfehler abzuschätzen, werden wir ein nichtlineares Céa-lemma herleiten, welches zeigt, dass der Diskretisierungsfehler mit dem bestmöglichen Fehler abgeschätzt werden kann. Um den bestmöglichen Fehler abzuschätzen, verallgemeinern wir eine Klasse von Approximationsoperatoren in Finite-Elemente-Räume und zeigen, dass die Verallgemeinerung die gleichen Fehlerabschätzungen erfüllt wie das lineare Gegenstück.

Zusammenfassend kann man sagen, dass die Arbeit numerische Methoden und Theorien für Optimierungs- und Variationsprobleme vom reellwertigen auf den mannigfaltigkeitswertigen Fall verallgemeinert.

Contents

Introduction

Many problems in physics and related disciplines can be formulated as optimization or variational problems. Sometimes the solution we seek has to satisfy nonlinear constraints. In liquid crystal physics [5] or micromagnetics [20] we seek vector fields with the constraint that the vectors have length 1, i.e. the vector field is a map from a domain $\Omega \subset \mathbb{R}^s, s \in \mathbb{N}$ into the sphere $S^n := \{x \in \mathbb{R}^{n+1} \mid |x| = 1\}$ with $|\cdot|$ the Euclidean norm. The sphere is a classical example of a Riemannian manifold. In this thesis, we design and analyze numerical methods for optimization and variational problems where we have the constraint that our data has to lie on a Riemannian manifold M. In an optimization problem, we want to find a minimizer of a functional $J\colon M \to \mathbb{R}$. In a variational problem, we want to find a minimizer of a functional $\mathcal{J}\colon H \to \mathbb{R}$ where H is a set of functions mapping Ω into M.

In Chapter 1, we introduce some basic concepts to deal with manifold-valued data. That includes Riemannian manifolds, geodesics, the exponential map, the closest point projection, optimization of manifold-valued functions and averages of manifold-valued data.

Optimization problems on Riemannian manifolds have already been studied in Absil et al. [3]. There, many optimization algorithms for functionals on the Euclidean space $\mathbb{R}^n$ are generalized to optimization algorithms for functionals on a Riemannian manifold M. The classical Newton method given by the iteration

$$\phi_{\mathbb{R}^n}(p) = p - (\operatorname{Hess} J(p))^{-1} \operatorname{grad} J(p)$$

can for example be generalized using the iteration

$$\phi_M(p) = \exp_p(-(\operatorname{Hess} J(p))^{-1} \operatorname{grad} J(p)), \tag{0.1}$$

where $\exp_p\colon T_pM \to M$ is the exponential map (Definition 1.1.4), T_pM the tangent space at $p \in M$, and $\operatorname{Hess} J(p)\colon T_pM \to T_pM$ and $\operatorname{grad} J(p) \in T_pM$ the Riemannian Hessian and gradient defined in Section 1.3.1.

In Chapter 2, we deal with a concrete example of an optimization problem. We consider images $V \to M$ where V is a set of pixels (usually a two-dimensional grid). Such images appear naturally in various signal and image processing applications. Some examples are:

- Grayscale images $V \to \mathbb{R}$ [12].

- RGB-images $V \to \mathbb{R}^3$ [31].
- Chromaticity components $V \to S^2$ [50] of RGB-images $u\colon V \to \mathbb{R}^3$ defined by $i \mapsto u_i/|u_i|$ for all $i \in V$.
- DT-MRI (diffusion tensor magnetic resonance images) $V \to \mathit{SPD}(3)$ [29], where $\mathit{SPD}(3)$ denotes the set of positive definite 3×3 matrices.

In many applications we are given only noisy measurements. Additionally, often various pixel values are corrupted which leaves us with noisy measurements on a subset $V_k \subset V$ of the pixel set. The task is to restore the image $u\colon V \to M$ from partial and noisy data $u^n\colon V_k \to M$ such that natural invariances of the manifold M are preserved. To solve this task we minimize the functional $J\colon M^V \to \mathbb{R}$ defined by

$$J(u) := \sum_{i \in V_k \subset V} d^2(u_i, u_i^n) + \lambda \sum_{(i,j)\in E} d(u_i, u_j) \text{ for all } u = (u_i)_{i\in V} \in M^V. \tag{0.2}$$

where $\lambda > 0$ is a positive constant balancing the fidelity and the total variation part of the functional, $E \subset V \times V$ is a given set of pairs of pixels that are considered to be close to each other and $d\colon M \times M \to \mathbb{R}_{\geq 0}$ is a metric on M. If V is a two-dimensional grid the edges $E \subset V \times V$ could be for example all pairs of pixels which are adjacent (horizontally or vertically).

Unfortunately, the functional (0.2) does not have the required amount of smoothness to apply the generalized Newton method (0.1) directly. To circumvent this problem, we define a series of smooth optimization problems with functionals of the form

$$J_w(u) := \sum_{i \in V_k \subset V} d^2(u_i, u_i^n) + \lambda \sum_{(i,j)\in E} w_{i,j} d^2(u_i, u_j) \text{ for all } u = (u_i)_{i\in V} \in M^V, \tag{0.3}$$

where the weights $w = (w_{i,j})_{(i,j)\in E} \subset \mathbb{R}_{>0}$ depend on the solution of the preceding problem. Note that in (0.3) all distances are squared whereas in (0.2) also distances without a square occur. To minimize J_w we propose to use the Riemannian Newton method (0.1). We call the resulting procedure the iteratively reweighted minimization (IRM) algorithm.

To study convergence properties of IRM, we examine under which conditions the functional J_w has a unique critical point and consequently also a unique minimizer. In the case $M = \mathbb{R}$ and d the Euclidean metric the functionals J respectively J_w are convex, respectively strictly convex. One can use this to prove the existence of a unique critical point of J_w. The same statement can be made for manifolds with non positive (sectional) curvature, the so-called Hadamard manifolds. However, there are manifolds where J is not convex and has multiple minimizers. An important example is the sphere (Example 2.4.1). However, if we restrict ourselves to the open half-sphere, we can, in spite of the non-convexity, prove uniqueness of a minimizer of J_w (Theorem 2.4.2). The idea is to prove local convexity at critical points and then use a tool from differential topology, namely the Poincaré–Hopf theorem. An interesting open problem is whether this result

for the half-sphere can be generalized to arbitrary manifolds, i.e. if for any Riemannian manifold M, and a geodesically closed subset $U \subset M$ homeomorphic to a ball the functional J_w restricted to U, defined in (0.3), with $u^n \colon V_k \to U$, has a unique critical point in U.

In Chapter 4, the goal is to numerically solve the following variational problem: given a functional $\mathcal{J} \colon H \to \mathbb{R}$, where H is a set of weakly differentiable functions from $\Omega \subset \mathbb{R}^s$ to a Riemannian manifold M, we want to find

$$u := \operatorname*{arg\,min}_{w \in H} \mathcal{J}(w). \tag{0.4}$$

A prototypical functional is the harmonic energy defined by

$$\mathcal{J}(u) := |u|^2_{H^1(\Omega,M)} := \int_\Omega \sum_{i=1}^{s} |\partial_i u(x)|^2_{g(u(x))} dx, \tag{0.5}$$

where ∂_i denotes the derivative in the i-th direction and $|\cdot|_{g(p)}$ is the norm on the tangent space T_pM. The set H could for example be all functions in $H^1(\Omega, M)$, subject to Dirichlet boundary conditions $u|_{\delta\Omega} = g \colon \delta\Omega \to M$. We propose and analyze two numerical methods to solve such problems.

The first method is called finite distance method and is inspired by the well-known finite difference method. The idea in the one-dimensional case is to approximate the integral of the squared norm of the derivative by the squared distance, i.e.

$$\int_x^{x+h} |\partial_i u(t)|^2 dt \approx h^{-1} d^2(u(x+h), u(x)).$$

The corresponding discretized functional is of the same form as (0.3) and we can apply the results for that problem (for example Theorem 2.4.2). We will call the corresponding algorithm the finite distance method.

The second method, which we will call geometric finite element method, is inspired by classical finite element theory. The idea is to solve the variational problem (0.4) on a subspace $V \subset H$, i.e. we seek

$$v := \operatorname*{arg\,min}_{w \in V} \mathcal{J}(w). \tag{0.6}$$

In the linear theory (i.e. when $M = \mathbb{R}$) the subspace $V_\mathbb{R}$ is usually a finite dimensional vector space, i.e.

$$V_\mathbb{R} = \left\{ \sum_{i \in I} p_i \phi_i \;\middle|\; p_i \in \mathbb{R} \right\}$$

where I is a finite index set, $\phi_i \colon \Omega \to \mathbb{R}$ a compactly supported function for all $i \in I$ and $(\phi_i)_{i \in I}$ a basis of $V_\mathbb{R}$. Assuming that $(\phi_i)_{i \in I}$ is a partition of unity, i.e. that $\sum_{i \in I} \phi_i(x) = 1$ holds for all $x \in \Omega$, we have that $\sum_{i \in I} p_i \phi_i(x)$ is a weighted average of the values $(p_i)_{i \in I} \subset \mathbb{R}$ with weights $(\phi_i(x))_{i \in I}$ for all $x \in \Omega$. In Chapter 1, we introduce

weighted averages for manifold-valued data $(p_i)_{i\in I} \subset M$. It follows that we can define a space V by replacing the linear combination with a weighted average av, i.e.

$$V := \{v\colon \Omega \to M,\ v(x) := av\left((\phi_i(x))_{i\in I}, (p_i)_{i\in I}\right)\ \mid p_i \in M \text{ for all } i \in I\}. \tag{0.7}$$

Note that a function in V is uniquely determined by the data $(p_i)_{i\in I} \subset M$. Hence problem (0.6) is an optimization problem on the Cartesian power M^I of the manifold M.

To estimate the error $|u-v|_{H^1}$, we prove a generalization (Lemma 4.3.1) of the classical Céa lemma, i.e. we show that for elliptic $\mathcal{J}$ (Definition 4.1.1) the discretization error is bounded by a constant times the best approximation error, i.e.

$$|u-v|_{H^1} \leq C \inf_{w\in V} |u-w|_{H^1},$$

where $C > 0$ is a constant depending only on the ellipticity constants of $\mathcal{J}\colon H \to \mathbb{R}$. It is then sufficient to study the approximation properties of the space V_M, which we will do in Chapter 3.

Notation

Before we start, we fix some notation we are going to use throughout the thesis.

We write A^T for the transpose of a matrix A. We denote the standard inner product by $\langle\cdot,\cdot\rangle$. For orthogonal $x, y \in \mathbb{R}^n$ (i.e. when $\langle x, y\rangle = 0$) we write $x \perp y$. By id we denote the identity function.

If for two quantities A and B depending on certain parameters there exists a constant $C > 0$ such that $A \leq CB$ independent of the choice of the parameters we write $A \lesssim B$.

If $(V, |\cdot|)$ and $(W, |\cdot|)$ are two normed vector spaces, $r \in \mathbb{N}$ and $A\colon V \to W$ satisfies $|A(v)| \lesssim |v|^r$ for $|v|$ sufficiently small we write $A(v) = \mathcal{O}(|v|^r)$. Similarly, if $|A(v)| < \epsilon|v|^r$ for any $\epsilon > 0$ and $|v|$ sufficiently small we write $A(v) = o(|v|^r)$.

For a function G and $r \in \mathbb{N}$ we denote by $G^{(r)}$ its r-th derivative. We sometimes also write G', G'' respectively G''' for the first, second respectively third derivative of G. For the r-th derivative at x applied to $y_1, y_2, \ldots, y_r$ we write $G^{(r)}(x)[y_1, \ldots, y_r]$. By $|\cdot|$ we denote the Euclidean norm and by $\|\cdot\|$ the operator norm with respect to the Euclidean norm. We write $C(X, Y)$ respectively $C^r(X, Y)$ for the space of all continuous respectively r-times continuously differentiable functions from X to Y. For $G \in C^r(X, Y)$ we define $|G|_{C^r} := \sup_{x\in X} \|G^{(r)}(x)\|$.

For $l \in \mathbb{N}$ and $p \in [1, \infty]$ we denote by $W^{l,p}$ the Sobolev space of l times weakly differentiable functions from a domain $\Omega \subset \mathbb{R}^s$ to $\mathbb{R}$ with derivatives in $L^p(\Omega)$. By $W^{l,p}(\Omega, \mathbb{R}^n)$

we denote the set of all measurable functions $v\colon \Omega \to \mathbb{R}^n$ such that $|v|_2 \in W^{l,p}$ where $|v|_2(x) := |v(x)|$ for all $x \in \Omega$. We define a seminorm and norm on $W^{l,p}(\Omega, \mathbb{R}^n)$ by

$$|v|_{W^{l,p}} := \left(\sum_{\substack{\vec{a} \in \mathbb{N}^s \\ |\vec{a}|_1 = l}} \left\| |D^{\vec{a}} v|_2 \right\|_{L^p}^p \right)^{\frac{1}{p}} \quad \text{and} \quad \|v\|_{W^{l,p}} := \left(\sum_{\substack{\vec{a} \in \mathbb{N}^s \\ |\vec{a}|_1 \leq l}} \left\| |D^{\vec{a}} v|_2 \right\|_{L^p}^p \right)^{\frac{1}{p}}, \tag{0.8}$$

where for $\vec{a} = (a_1, \ldots, a_s) \in \mathbb{N}^s$ we define $|\vec{a}|_1 := \sum_{i=1}^s a_i$ and

$$D^{\vec{a}} := \frac{\partial^{|\vec{a}|_1}}{\partial x_1^{a_1} \ldots \partial x_s^{a_s}}. \tag{0.9}$$

We will write H^l for $W^{l,2}$. Note that (0.8) with $l = 1$ and $p = 2$ is compatible with (0.5).

1 Preliminaries

In this chapter, we develop the mathematical basics to deal with manifold-valued data. A major difficulty is that a priori addition, scalar multiplication and more generally linear combinations are not defined for manifold-valued data. This makes it difficult or impossible to apply tools from the linear theory (i.e. when $M = \mathbb{R}$). To partially overcome this issue, we use weighted averages of points on a Riemannian manifold. First, we will have to study some elementary Riemannian geometry. As we will see in Section 1.1, there is a natural way of defining an "addition" of a tangent vector $v \in T_pM$ to its base point $p \in M$ by the exponential map $\exp_p$ (Definition 1.1.4). Since in most applications our data lies on a Riemannian submanifold of $\mathbb{R}^n$, we will pay special attention to this case. [1] This will be done in Section 1.2, where we introduce an "addition" of a tangent vector $v \in T_pM$ to its base point $p \in M$ which, in most cases, is easier to implement than the exponential map. We show that this "addition" is numerically close to the exponential map (Proposition 1.2.8). As a preparation for the minimization of functionals $J\colon M^N \to \mathbb{R},\ N \in \mathbb{N}$ we introduce in Section 1.3 the Riemannian gradient, the Riemannian Hessian, the Taylor expansion of functions on manifolds and the Riemannian Newton method. In Section 1.4, we define weighted averages of manifold-valued data. Finally, in Section 1.5, we take a look at some specific manifolds one often encounters in real-world applications.

1.1 Riemannian geometry

In this section, we introduce the necessary concepts of Riemannian geometry. We assume that the reader is familiar with the basic notions of differentiable manifolds, tangent spaces and vector fields on manifolds.

1.1.1 Riemannian manifolds

We start with the definition of Riemannian manifolds. This additional structure on a manifold will enable us to measure distances.

[1] As every Riemannian manifold can be isometrically embedded in $\mathbb{R}^n$ for some $n \in \mathbb{N}$ large enough [39] one might think that we could restrict ourselves to Riemannian submanifolds of $\mathbb{R}^n$. However, if the embedding is not known it is not possible to use this fact in practice.

Definition 1.1.1. A *Riemannian manifold* is a differentiable manifold together with a family of positive definite inner products $g_p \colon T_pM \times T_pM \to \mathbb{R}$, $p \in M$ on the tangent spaces T_pM, such that for all differentiable vector fields X, Y on M the map $p \mapsto g_p(X(p), Y(p))$ is a smooth (i.e. C^∞) function.

1.1.2 The geodesic distance and geodesics

The inner product g_p on T_pM induces a norm on T_pM defined by $|v| := \sqrt{g_p(v,v)}$ for $v \in T_pM$. Using this norm we can define the geodesic distance on M.

Definition 1.1.2. The *geodesic distance* $d_g \colon M \times M \to \mathbb{R}$ on a Riemannian manifold M is defined by

$$d_g(p,q) := \inf_{\substack{\gamma \in C^1([0,1],M) \\ \gamma(0)=p, \gamma(1)=q}} \int_0^1 |\dot\gamma(t)| dt,$$

where $\dot\gamma(t) \in T_{\gamma(t)}M$ is the derivative of γ at $t \in \mathbb{R}$. A curve $\gamma \in C^1([a,b], M)$ is called *length-minimizing* if $d_g(\gamma(a), \gamma(b)) = \int_a^b |\dot\gamma(t)| dt$. A curve $\gamma \in C^1(I, M)$, where $I \subset \mathbb{R}$ is an interval, is called a *geodesic* if it is locally length-minimizing and of constant speed, i.e. if there exists $s \geq 0$ such that for every $x \in \mathbb{R}$ there exists $\epsilon > 0$ such that for all $y \in \mathbb{R}$ with $|x - y| < \epsilon$ we have

$$d_g(\gamma(x), \gamma(y)) = |y - x| s. \tag{1.1}$$

A length-minimizing geodesic can also be expressed as the minimum of an energy functional:

Proposition 1.1.3. *Let $p, q \in M$ and $\gamma \in C^1([0,1], M)$ with $\gamma(0) = p$, $\gamma(1) = q$. Then γ is a length-minimizing geodesic if and only if*

$$\gamma \in \operatorname*{arg\,min}_{\substack{\alpha \in C^1([0,1],M) \\ \alpha(0)=p, \alpha(1)=q}} \int_0^1 |\dot\alpha(t)|^2 dt. \tag{1.2}$$

Proof. By the Cauchy–Schwarz inequality and the definition of the geodesic distance we have

$$\int_0^1 |\dot\gamma(t)|^2 dt = \left(\int_0^1 1^2 dt \right) \int_0^1 |\dot\gamma(t)|^2 dt \geq \left(\int_0^1 |\dot\gamma(t)| dt \right)^2 \geq d_g^2(p,q)$$

with equality if and only if γ is a length-minimizing curve with constant speed, i.e. if and only if γ is a length-minimizing geodesic. □

1.1.3 The exponential map

In this section, we define the exponential map. It can be seen as a way to "add" a tangent vector $v \in T_pM$ to its base point $p \in M$. By proving that the corresponding ordinary differential equation has a unique solution it can be shown that for $p \in M$ and $v \in T_pM$ there exists a unique geodesic $\gamma \in C^1(\mathbb{R}, M)$ with $\gamma(0) = p$ and $\dot{\gamma}(0) = v$.

Definition 1.1.4. The *exponential map* $exp_p \colon T_pM \to M$ *at* $p \in M$ is defined by $\exp_p(v) := \gamma_v(1)$ for every $v \in T_pM$ where $\gamma_v \in C^1(\mathbb{R}, M)$ is the geodesic with $\gamma_v(0) = p$ and $\dot{\gamma}_v(0) = v$.

It can be shown that the exponential map at $p \in M$ is a local diffeomorphism from a neighborhood of $\{0\}$ in T_pM onto a neighborhood U_p of p in M. Hence, $\exp_p$ is locally invertible. The inverse of $\exp_p$ is called the *logarithm map at* $p \in M$ and it is denoted by $\log_p$. The function $\log_p \colon U_p \to T_pM$ can be thought of as a way to "subtract" $p \in M$ from another point on the manifold.

1.2 Riemannian submanifolds of $\mathbb{R}^n$

In this section, we consider the case where M is a submanifold of $\mathbb{R}^n$. Note that for a submanifold $M \subset \mathbb{R}^n$ there is a natural embedding $T_pM \hookrightarrow \mathbb{R}^n$ for all $p \in M$. If we equip a submanifold of $\mathbb{R}^n$ with the standard inner product $\langle x, y \rangle := \sum_{i=1}^n x_i y_i$ we get a Riemannian manifold.

Definition 1.2.1. A *Riemannian submanifold* M *of* $\mathbb{R}^n$ is a submanifold of $\mathbb{R}^n$ equipped with the standard inner product.

An important tool for us is the closest point projection $\mathcal{P}$ which simply maps a point in $\mathbb{R}^n$ to its nearest point on M.

Definition 1.2.2. For a submanifold M of $\mathbb{R}^n$ the *closest point projection* $\mathcal{P}(p)$ *of* $p \in \mathbb{R}^n$ *onto* M is defined by

$$\mathcal{P}(p) := \operatorname*{arg\,min}_{q \in M} |q - p|.$$

The closest point projection $\mathcal{P}$ is usually not well-defined for all $p \in \mathbb{R}^n$. For the sphere $S^{n-1} := \{x \in \mathbb{R}^n \mid |x| = 1\} \subset \mathbb{R}^n$ the closest point projection $\mathcal{P}$ is for example not defined at $p = 0$. However, if M is sufficiently regular the closest point projection is well-defined in a neighborhood of M [1]. We denote this neighborhood by U. Using the closest point projection $\mathcal{P} \colon U \subset \mathbb{R}^n \to M$ we can define a second way of adding a tangent vector $v \in T_pM$ to its base point p, which we call the projection-based retraction. Because of the natural embedding $T_pM \hookrightarrow \mathbb{R}^n$ the sum $p + v \in \mathbb{R}^n$ is well-defined for any $p \in M$ and $v \in T_pM$. To get back to the manifold we compose this sum with the closest point projection $\mathcal{P} \colon U \subset \mathbb{R}^n \to M$.

Definition 1.2.3. Let $M \subset \mathbb{R}^n$ be a submanifold and $\mathcal{P}\colon U \subset \mathbb{R}^n \to M$ the closest point projection onto M. Then the *projection-based retraction* $e_p\colon T_pM \cap (U - p) \to M$ *at* $p \in M$ is defined by $e_p(v) := \mathcal{P}(p + v)$ for all $v \in T_pM \cap (U - p)$, where

$$U - p := \{u - p \mid u \in U\}.$$

Together with the exponential map we now have two ways of adding a tangent vector to a basepoint $p \in M$.

In Section 1.2.1, we show that the derivative of the closest point projection at a point on M is the orthogonal projection onto the tangent space. In Section 1.2.2 we give some connections of geodesics on Riemannian submanifolds of $\mathbb{R}^n$ and the closest point projection $\mathcal{P}$. This will be applied in Section 1.2.3 where we estimate the difference between the exponential map and the projection-based retraction.

1.2.1 Derivative of the closest point projection $\mathcal{P}$

The following property of the closest point projection $\mathcal{P}$ will turn out to be useful.

Lemma 1.2.4. *Let $M \subset \mathbb{R}^n$ be a submanifold. Assume that the closest point projection $\mathcal{P}\colon U \subset \mathbb{R}^n \to M$ onto M is differentiable on M. Then $\mathcal{P}'(p)\colon \mathbb{R}^n \to T_pM$ is the orthogonal projection onto the tangent space T_pM for all $p \in M$.*

Proof. Let $p \in M$ and $\gamma \in C^1([0,1], M)$ with $\gamma(0) = p$. Since $\mathcal{P}(\gamma(t)) = \gamma(t)$ for all $t \in [0,1]$ we have by differentiating that $\mathcal{P}'(\gamma(0))[\dot\gamma(0)] = \dot\gamma(0)$. Hence, $\mathcal{P}'(p)$ restricted to the tangent space is the identity. Let $v \in T_pM^\perp$ where $T_pM^\perp$ is the normal space at $p \in M$, i.e. the orthogonal complement of T_pM. Since $\mathcal{P}$ is the closest point projection we have

$$|\mathcal{P}(p + tv) - (p + tv)| \leq |p - (p + tv)| = |tv| = |t||v|$$

whenever $\mathcal{P}(p+tv)$ is well-defined. Let $w := \mathcal{P}'(p)[v]$. Since $w \in T_pM$ we have $\langle w, v\rangle = 0$ and using Taylor expansion

$$\begin{aligned} |t||v| &\geq |\mathcal{P}(p + tv) - (p + tv)| \\ &= |p + t\mathcal{P}'(p)[v] + o(|t|) - (p + tv)| \\ &= |t(w - v) + o(|t|)| \\ &= |t|\sqrt{|v|^2 + |w|^2} + o(|t|), \end{aligned}$$

and therefore

$$|t|\left(\sqrt{|v|^2 + |w|^2} - |v|\right) = o(|t|), \quad \sqrt{|v|^2 + |w|^2} = |v|$$

and hence $\mathcal{P}'(p)[v] = w = 0$. □

1.2.2 Geodesics on Riemannian submanifolds of $\mathbb{R}^n$

For a Riemannian submanifold M of $\mathbb{R}^n$ we can characterize the geodesics $\gamma \in C^1(\mathbb{R}, M)$ more precisely. We first show that $\ddot{\gamma}(t) \in T_{\gamma(t)}M^\perp$, i.e. the second derivative of γ is orthogonal to the tangent space.

Proposition 1.2.5. *Let M be a Riemannian submanifold of $\mathbb{R}^n$ and $\gamma \in C^2([a,b], M)$ a geodesic. Then we have $\ddot{\gamma}(t) \in T_{\gamma(t)}M^\perp$ for all $t \in (a,b)$.*

Proof. Let $w\colon [a,b] \to \mathbb{R}^n$ be a function with $w(t) \in T_{\gamma(t)}M$ for all $t \in (a,b)$ and $w(a) = w(b) = 0$. Define $\gamma^\epsilon\colon [a,b] \to M$ by $\gamma^\epsilon(t) = \mathcal{P}(\gamma(t) + \epsilon w(t))$. As γ is a critical point of the energy functional (1.2) we have using Lemma 1.2.4 and integration by parts that

$$0 = \frac{d}{d\epsilon}\langle \dot{\gamma^\epsilon}, \dot{\gamma^\epsilon}\rangle_{L^2}|_{\epsilon=0} = 2\left\langle \frac{d}{d\epsilon}\dot{\gamma^\epsilon}|_{\epsilon=0}, \dot{\gamma}\right\rangle_{L^2} = 2\langle \dot{w}, \dot{\gamma}\rangle_{L^2} = -2\langle w, \ddot{\gamma}\rangle_{L^2}.$$

As this equation holds for all $w\colon [a,b] \to \mathbb{R}^n$ with $w(t) \in T_{\gamma(t)}M$ for all $t \in (a,b)$ and $w(a) = w(b) = 0$, we have $\ddot{\gamma}(t) \in T_{\gamma(t)}M^\perp$ for all $t \in (a,b)$. □

The following lemma relates the second derivative of a geodesic with its first derivative. The equation is of the same form as the classical geodesic equation of Riemannian geometry. From now on we will assume some regularity on the closest point projection $\mathcal{P}$ or equivalently on the submanifold $M \subset \mathbb{R}^n$.

Proposition 1.2.6. *Let M be a Riemannian submanifold of $\mathbb{R}^n$ and $\gamma \in C^2(\mathbb{R}, M)$ be a geodesic. Assume that the closest point projection $\mathcal{P}\colon \mathbb{R}^n \to M$ onto M is two times differentiable on M. Then we have*

$$\ddot{\gamma}(t) = \mathcal{P}''(\gamma(t))[\dot{\gamma}(t), \dot{\gamma}(t)] \text{ for all } t \in \mathbb{R}. \tag{1.3}$$

Proof. As γ takes values on M and $\mathcal{P}$ is a projection onto M we have $\gamma(t) = \mathcal{P}(\gamma(t))$. Taking two derivatives with respect to t using the chain rule yields

$$\ddot{\gamma}(t) = \mathcal{P}'(\gamma(t))[\ddot{\gamma}(t)] + \mathcal{P}''(\gamma(t))[\dot{\gamma}(t), \dot{\gamma}(t)].$$

By Lemma 1.2.4 and Proposition 1.2.5 the first term on the right hand side vanishes and we get (1.3). □

An immediate consequence of Proposition 1.2.6 is that the norm of the second derivative of a geodesic can be controlled by its first derivative assuming that the second derivative of the closest point projection $\mathcal{P}$ is bounded. This is also equivalent to the radius of curvature of M being bounded.

Corollary 1.2.7. *Consider the same situation as in Proposition 1.2.6. Assume that the second derivative of the closest point projection $\mathcal{P}$ onto M is bounded on M. Then we have*

$$|\ddot{\gamma}(t)| \leq |\mathcal{P}|_{C^2}|\dot{\gamma}(t)|^2.$$

1.2.3 Proximity of the exponential map and the projection-based retraction

In this section, we show that the projection-based retraction e_p is close to the exponential map $\exp_p$. The projection-based retraction is a retraction of order 2, i.e. we have $e_p(0) = p = \exp_p(0)$, $e_p'(0) = id = \exp_p'(0)$ and $e_p''(0) = \exp_p''(0)$. This allows us to estimate $|\exp_p(v) - e_p(v)|$ by $|v|^3$. A more general statement can be found in [2].

Proposition 1.2.8. *Let M be a Riemannian submanifold of $\mathbb{R}^n$ and $C > 0$. Assume that the closest point projection $\mathcal{P}\colon U \subset \mathbb{R}^n \to M$ onto M is three times differentiable on $A := \{p + v \mid p \in M, v \in T_pM \cap (U - p) \text{ with } |v| < C\}$ and the derivatives can be uniformly bounded. Then we have*

$$|exp_p(v) - e_p(v)| \lesssim |v|^3 \text{ for all } p \in M \text{ and } v \in T_pM \text{ with } |v| < C,$$

where the implicit constant depends only on the bounds for the first three derivatives of $\mathcal{P}$ on A.

Proof. Let $p \in M$ and $v \in \cap(U-p)$ with $|v| < C$. Consider the geodesic $\gamma_g(t) := \exp_p(tv)$ and the curve $\gamma_p(t) := e_p(tv)$. We have $\gamma_g(0) = p = \gamma_p(0)$, $\dot\gamma_g(0) = v = \dot\gamma_p(0)$ and by Proposition 1.2.6 $\ddot\gamma_g(0) = \mathcal{P}''(p)[v, v] = \ddot\gamma_p(0)$. Hence we have by Proposition 1.2.6 and Corollary 1.2.7

$$\begin{aligned}
|\gamma_g(1) - \gamma_p(1)| &= \left|\frac{1}{2}\int_0^1 (1-t)^2 \left(\gamma_g^{(3)}(t) - \gamma_p^{(3)}(t)\right) dt\right| \\
&\leq \frac{1}{2}\int_0^1 \left|\gamma_g^{(3)}(t)\right| + \left|\gamma_p^{(3)}(t)\right| dt \\
&\leq \frac{1}{2}\int_0^1 |\mathcal{P}'''(\gamma_g(t))[\dot\gamma_g(t), \dot\gamma_g(t), \dot\gamma_g(t)] + 2\mathcal{P}''(\gamma_g(t))[\dot\gamma_g(t), \ddot\gamma_g(t)]| \\
&\quad + \left|\mathcal{P}'''(p + tv)[v, v, v]\right| dt \\
&\lesssim \int_0^1 |\dot\gamma_g(t)|^3 + |\dot\gamma_g(t)|\,|\ddot\gamma_g(t)| + |v|^3 dt \\
&\lesssim \int_0^1 |\dot\gamma_g(t)|^3 + |v|^3 dt \\
&\lesssim |v|^3. \quad \square
\end{aligned}$$

In Proposition 1.2.8, we used the Euclidean distance $d_E(p, q) := |p - q|$ to measure the distance between $\exp_p(v)$ and $e_p(v)$. One might ask what would happen if we would use the geodesic distance d_g of Definition 1.1.2 instead. The following lemma shows that for points close to each other the Euclidean distance is a good approximation of the geodesic distance. Proposition 1.2.8 would also hold if the Euclidean distance is replaced by the geodesic distance.

Lemma 1.2.9. *Let M be a Riemannian submanifold of $\mathbb{R}^n$. Assume that the closest point projection $\mathcal{P}\colon \mathbb{R}^n \to M$ onto M is two times differentiable on M. Then we have*

$$|log_p(q) - (q-p)| \leq \frac{1}{2}|\mathcal{P}|_{C^2} d_g^2(p,q).$$

and

$$0 \leq d_g(p,q) - d_E(p,q) \leq \frac{1}{2}|\mathcal{P}|_{C^2}^2 d_g^3(p,q).$$

Proof. Let $\gamma\colon \mathbb{R} \to M$ be a geodesic with $\gamma(0) = p$, $\gamma(1) = q$ and $d_g(p,q) = |\dot\gamma(0)|$. By Corollary 1.2.7 we have

$$|\log_p(q) - (q-p)| = |\dot\gamma(0) - (\gamma(1) - \gamma(0))| = \left|\dot\gamma(0) - \int_0^1 \dot\gamma(t)dt\right|$$

$$= \left|\int_0^1 (1-t)\ddot\gamma(t)dt\right| \leq \int_0^1 (1-t)|\mathcal{P}|_{C^2}|\dot\gamma(t)|^2 dt = \frac{1}{2}|\mathcal{P}|_{C^2} d_g^2(p,q).$$

As the straight line is the closest connection between two points in a Euclidean space we have $d_E(p,q) \leq d_g(p,q)$. To prove the other direction note that we have

$$\begin{aligned}
\langle\dot\gamma(0), \gamma(1) - \gamma(0)\rangle &= \langle\dot\gamma(0), \dot\gamma(0)\rangle - \left\langle\dot\gamma(0), \int_0^1 (1-t)\ddot\gamma(t)\right\rangle \\
&= |\dot\gamma(0)|^2 + \int_0^1 (1-t)\langle\dot\gamma(0), \ddot\gamma(t)\rangle dt \\
&= |\dot\gamma(0)|^2 + \int_0^1 (1-t)\left(\langle\dot\gamma(t), \ddot\gamma(t)\rangle - \int_0^t \langle\ddot\gamma(s), \ddot\gamma(t)\rangle ds\right) dt \\
&= |\dot\gamma(0)|^2 - \int_0^1 (1-t)\int_0^t \langle\ddot\gamma(s), \ddot\gamma(t)\rangle ds dt \\
&\geq d_g^2(p,q) - \frac{1}{2}|\mathcal{P}|_{C^2}^2 d_g^4(p,q)
\end{aligned}$$

where we used $\dot\gamma(t) \perp \ddot\gamma(t)$ and Corollary 1.2.7. Therefore by the Cauchy–Schwarz inequality

$$\begin{aligned}
d_g(p,q) - d_E(p,q) &= \frac{d_g^2(p,q) - |\dot\gamma(0)||\gamma(1) - \gamma(0)|}{d_g(p,q)} \\
&\leq \frac{d_g^2(p,q) - \langle\dot\gamma(0), \gamma(1) - \gamma(0)\rangle}{d_g(p,q)} \\
&\leq \frac{1}{2}|\mathcal{P}|_{C^2}^2 d_g^3(p,q). \quad \square
\end{aligned}$$

1.3 Optimization of functionals on manifolds

In Chapter 2 and 4, we are in a situation in which we have to find minimizers of functionals $J\colon M \to \mathbb{R}$. As a preparation, we introduce the Riemannian gradient and Riemannian Hessian of such functionals J in this section. The Riemannian Hessian is usually introduced using the covariant derivative. This requires the knowledge of some advanced differential geometry. However, the Riemannian Hessian of a functional J at $p \in M$ is also related with the classical Hessian by a simple relation involving the exponential map (see, e.g. [3]). We will use this relation to give an alternative definition of the Riemannian Hessian which makes the thesis accessible to readers unfamiliar with advanced differential geometry. To be able to minimize functionals on manifolds, we will, in Section 1.3.2, take a look at the generalization of the gradient descent and in Section 1.3.3 at the generalization of the Newton method.

1.3.1 Gradient, Hessian and Taylor expansion on manifolds

Let $J\colon M \to \mathbb{R}$ be a two times differentiable functional on a Riemannian manifold M. Note that the composition $J \circ \exp_p\colon T_pM \to \mathbb{R}$ is then a two times differentiable map from the vector space T_pM into $\mathbb{R}$. Hence, there exists $\operatorname{grad} J(p) \in T_pM$ and a self-adjoint operator $\operatorname{Hess} J(p)\colon T_pM \to T_pM$ of $J \circ \exp_p$ at $p \in M$ such that we have the Taylor expansion

$$J(\exp_p(v)) = J(p) + g_p(\operatorname{grad} J(p), v) + \frac{1}{2} g_p(v, \operatorname{Hess} J(p)v) + o(|v|^2), \tag{1.4}$$

where g_p is the inner product on T_pM (see Definition 1.1.1). The vector $\operatorname{grad} J(p) \in T_pM$ is called the Riemannian gradient of J. The operator $\operatorname{Hess} J(p)\colon T_pM \to T_pM$ is called the Riemannian Hessian of J. Note that for a Riemannian submanifold M of $\mathbb{R}^n$ we have by Proposition 1.2.8 that

$$J(e_p(v)) = J(p) + \langle \operatorname{grad} J(p), v\rangle + \frac{1}{2}\langle v, \operatorname{Hess} J(p)v\rangle + o(|v|^2), \tag{1.5}$$

where e_p is defined in Definition (1.2.3).

For some functionals $J\colon M \to \mathbb{R}$, with M being a Riemannian submanifold of $\mathbb{R}^n$, there exists a natural extension $\bar{J}\colon U \subset \mathbb{R}^n \to \mathbb{R}$ (or J is actually the restriction of $\bar{J}$ to the manifold), where U is a neighborhood of M. In some cases it is relatively simple to compute the classical gradient $\operatorname{grad} \bar{J}(p) \in \mathbb{R}^n$ and Hessian $\operatorname{Hess} \bar{J}(p) \in \mathbb{R}^{n\times n}$ at $p \in M$ of $\bar{J}$. The next proposition gives relations between the Riemannian gradient $\operatorname{grad} J(p)$ respectively Hessian $\operatorname{Hess} J(p)$ and the classical gradient respectively Hessian.

Proposition 1.3.1. *Let M be a Riemannian submanifold of $\mathbb{R}^n$, $J\colon M \to \mathbb{R}$ a two times differentiable functional and $\bar{J}\colon U \subset \mathbb{R}^n \to \mathbb{R}$ a two times differentiable extension of J.*

Assume that the closest point projection $\mathcal{P}$ is two times differentiable on M. Then we have for $p \in M$ and $v \in T_pM$ that

$$\langle \operatorname{grad} J(p), v\rangle = \langle \operatorname{grad} \bar{J}(p), v\rangle$$

and

$$\langle v, \operatorname{Hess} J(p)v\rangle = \langle v, \operatorname{Hess} \bar{J}(p)v\rangle + \langle \operatorname{grad} \bar{J}(p), \mathcal{P}''(p)[v,v]\rangle. \tag{1.6}$$

Proof. Taylor expansion of $e_p(v) = \mathcal{P}(p+v)$ at 0 and of $\bar{J}$ at p yields

$$e_p(v) = p + v + \frac{1}{2}\mathcal{P}''(p)[v,v] + o(|v|^2) \tag{1.7}$$

and

$$\bar{J}(p+w) = \bar{J}(p) + \langle \operatorname{grad} \bar{J}(p), w\rangle + \frac{1}{2}\langle w, \operatorname{Hess} \bar{J}(p)w\rangle + o(|w|^2).$$

Choosing $w = e_p(v) - p = v + \frac{1}{2}\mathcal{P}''(p)[v,v] + o(|v|^2)$ yields

$$\begin{aligned}\bar{J}(p+w) &= \bar{J}(p) + \langle \operatorname{grad} \bar{J}(p), v + \frac{1}{2}\mathcal{P}''(p)[v,v]\rangle + \frac{1}{2}\langle v, \operatorname{Hess} \bar{J}(p)v\rangle + o(|v|^2)\\ &= \bar{J}(p) + \langle \operatorname{grad} \bar{J}(p), v\rangle + \frac{1}{2}\Big(\langle v, \operatorname{Hess} \bar{J}(p)v\rangle + \langle \operatorname{grad} \bar{J}(p), \mathcal{P}''(p)[v,v]\rangle\Big)\\ &\quad + o(|v|^2).\end{aligned}$$

Since $\bar{J}(p+w) = \bar{J}(e_p(v)) = J(e_p(v))$ the statement can be deduced by comparing with (1.5). □

An alternative approach for computing the Riemannian Hessian from the classical Hessian can be found in [4]. It requires however the Weingarten map.

In some situations it is possible to compute the gradient $\operatorname{grad} J\colon M \to \mathbb{R}^n$, with M being a Riemannian submanifold of $\mathbb{R}^n$ of a functional $J\colon M \to \mathbb{R}$ and find an extension $G\colon U \subset \mathbb{R}^n \to \mathbb{R}^n$ of $\operatorname{grad} J$. The following proposition gives a relation between the Hessian of J and the classical derivative of G.

Proposition 1.3.2. *Let M be a Riemannian submanifold of $\mathbb{R}^n$, $J\colon M \to \mathbb{R}$ a two times differentiable functional, $\operatorname{grad} J\colon M \to TM$ the gradient of J and $G\colon U \subset \mathbb{R}^n \to \mathbb{R}^n$ a differentiable extension of $\operatorname{grad} J$. Assume that the closest point projection $\mathcal{P}$ is two times differentiable on M. Then we have for $p \in M \subset \mathbb{R}^n$ and $v \in T_pM$ that*

$$\langle v, \operatorname{Hess} J(p)v\rangle = \langle v, G'(p)[v]\rangle. \tag{1.8}$$

Proof. Let $q := \exp_p(v)$ and $\gamma\colon [0,1] \to M$ the geodesic from p to q. Using $G(p) \in T_pM$ and $\ddot{\gamma}(0) \in T_pM^\perp$ we have $\langle G(p), \ddot{\gamma}(0)\rangle = 0$ and hence

$$J(\gamma(1)) - J(\gamma(0)) = \int_0^1 \frac{d}{dt} J(\gamma(t))dt$$

$$
\begin{aligned}
&= \int_0^1 \langle \operatorname{grad} J(\gamma(t)), \dot\gamma(t)\rangle dt \\
&= \int_0^1 \langle G(\gamma(t)), \dot\gamma(t)\rangle dt \\
&= \int_0^1 \langle G(\gamma(0)) + G'(\gamma(0))[\gamma(t)-\gamma(0)] + o(|v|), \dot\gamma(0) + t\ddot\gamma(0) + o(|v|^2)\rangle dt \\
&= \int_0^1 \langle G(p) + G'(p)[tv] + o(|v|\rangle, v + t\ddot\gamma(0) + o(|v|^2)\rangle dt \\
&= \int_0^1 \langle G(p), v\rangle + t\langle G'(p)[v], v\rangle + o(|v|^2) dt \\
&= \langle \operatorname{grad} J(p), v\rangle + \frac{1}{2}\langle v, G'(p)[v]\rangle + o(|v|^2).
\end{aligned}
$$

Comparing with (1.4) yields the desired result. □

The relations in Proposition 1.3.1 respectively 1.3.2 will allow us to compute so called representations for the Riemannian gradient and the Riemannian Hessian.

Definition 1.3.3. Let M be a Riemannian submanifold of $\mathbb{R}^n$. A *representation of the gradient* $\operatorname{grad} J(p) \in T_pM$*, respectively the Hessian* $\operatorname{Hess} J(p)\colon T_pM \to T_pM$*, of a two times differentiable functional* $J\colon M \to \mathbb{R}$ *at* $p \in M$ is a vector $g \in \mathbb{R}^n$ respectively a matrix $H \in \mathbb{R}^{n\times n}$ such that

$$\langle v, g\rangle = \langle v, \operatorname{grad} J(p)\rangle \quad \text{and} \quad \langle v, Hv\rangle = \langle v, \operatorname{Hess} J(p)v\rangle \text{ for all } v \in T_pM.$$

To evaluate Hv the vector $v \in T_pM \subset \mathbb{R}^n$ has to be regarded as a vector in $\mathbb{R}^n$.

Note that if $H \in \mathbb{R}^{n\times n}$ is a representation of $\operatorname{Hess} J(p)$ then $\operatorname{Sym}(H) := (H + H^T)/2$ is also a representation for $\operatorname{Hess} J(p)$. This new representation has the advantage of being symmetric. For symmetric representations H the equation $\langle u, Hv\rangle = \langle u, \operatorname{Hess} J(p)v\rangle$ does not only hold on the diagonal, as is required to be a representation, but for all $u, v \in T_pM$.

Proposition 1.3.4. *Let* M *be a Riemannian submanifold of* $\mathbb{R}^n$*,* $p \in M$*,* $J\colon M \to \mathbb{R}$ *and* $H \in \mathbb{R}^{n\times n}$ *a symmetric representation of* $\operatorname{Hess} J(p)$*. Then we have*

$$\langle u, Hv\rangle = \langle u, \operatorname{Hess} J(p)v\rangle \text{ for all } u, v \in T_pM.$$

Proof. Since H is by assumption and $\operatorname{Hess} J(p)$ by definition self-adjoint we have

$$
\begin{aligned}
4\langle u, Hv\rangle &= \langle u+v, H(u+v)\rangle - \langle u-v, H(u-v)\rangle \\
&= \langle u+v, \operatorname{Hess} J(p)(u+v)\rangle - \langle u-v, \operatorname{Hess} J(p)(u-v)\rangle \\
&= 4\langle u, \operatorname{Hess} J(p)v\rangle. \quad \square
\end{aligned}
$$

In Chapter 2 and 4, we will minimize functionals $J\colon M^k \to \mathbb{R}$ where $k \in \mathbb{N}$ and $M^k := \prod_{i=1}^k M$. The Cartesian power M^k is again a manifold and its tangent space T_pM^k at $p = (p_i)_{i=1}^k \in M^k$ is the Cartesian product of the tangent spaces of M at the points $(p_i)_{i=1}^k$, i.e. we have $T_pM^k = \prod_{i=1}^k T_{p_i}M$. With the natural inner product

$$g_p(v,w) := \sum_{i=1}^k g_{p_i}(v_i, w_i) \text{ for all } v = (v_i)_{i=1}^k, w = (w_i)_{i=1}^k \in T_pM^k$$

the manifold M^k becomes a Riemannian manifold. The gradient $\operatorname{grad} J(p) \in T_pM^k$ of a differentiable functional J at $p \in M^k$ is

$$\operatorname{grad} J(u) = (\operatorname{grad}_1 J(u), \dots, \operatorname{grad}_k J(u))$$

where $\operatorname{grad}_i J(u) \in T_{p_i}M$ is the gradient of the functional with respect to the i-th coordinate. If M is a Riemannian submanifold of $\mathbb{R}^n$ and $\bar{J}\colon U \subset (\mathbb{R}^n)^k \to \mathbb{R}$ an extension of J we can also consider the classical gradient $\operatorname{grad} \bar{J}(p) \in (\mathbb{R}^n)^k \sim \mathbb{R}^{nk}$ and Hessian $\operatorname{Hess} \bar{J}(p) \in (\mathbb{R}^{n\times n})^{k\times k} \sim \mathbb{R}^{nk\times nk}$ of $\bar{J}$. Corollary 1.3.5 gives a relation between the Riemannian gradient $\operatorname{grad} J(p)$ and Riemannian Hessian $\operatorname{Hess} J(p)$ with the classical gradient $\operatorname{grad} \bar{J}(p)$ and classical Hessian $\operatorname{Hess} \bar{J}(p)$ of an extension $\bar{J}\colon U \subset (\mathbb{R}^n)^k \to \mathbb{R}$ of J.

Corollary 1.3.5. *Let M be a Riemannian submanifold of $\mathbb{R}^n$, $k \in \mathbb{N}$, $J\colon M^k \to \mathbb{R}$ a two times differentiable functional and $\bar{J}\colon U \subset (\mathbb{R}^n)^k \to \mathbb{R}$ a two times differentiable extension of J. Assume that the closest point projection $\mathcal{P}$ is two times differentiable on M. Then we have for $p \in M^k$ and $v \in T_pM^k$ that*

$$\langle \operatorname{grad} J(p), v\rangle = \langle \operatorname{grad} \bar{J}(p), v\rangle$$

and

$$\langle v, \operatorname{Hess} J(p)v\rangle = \langle v, \operatorname{Hess} \bar{J}(p)v\rangle + \sum_{i=1}^k \langle \operatorname{grad}_i \bar{J}(p), \mathcal{P}''(p_i)[v_i, v_i]\rangle. \tag{1.9}$$

Proof. This follows from Proposition 1.3.1 together with the fact that the exponential map on M^k respectively the closest point projection onto M^k is the Cartesian product of the exponential map on M respectively the closest point projection onto M. □

1.3.2 Gradient descent

The gradient descent method is one of the simplest algorithms to numerically compute a minimum of a functional. The straight forward generalization to functionals defined on a manifold is given below.

Definition 1.3.6. For a Riemannian manifold M and a differentiable functional $J\colon M \to \mathbb{R}$ the *gradient descent method with step size* $t \in \mathbb{R}$ is given by the iteration

$$\phi_R(p) = \exp_p(-t \operatorname{grad} J(p)).$$

If $M \subset \mathbb{R}^n$ is a Riemannian submanifold of $\mathbb{R}^n$ the *projection-based gradient descent method with step size* $t \in \mathbb{R}$ is given by the iteration

$$\phi_{\mathcal{P}}(p) = e_p(-t \operatorname{grad} J(p)).$$

1.3.3 The Riemannian Newton method

The classical Newton method can be used to find a critical point (i.e. a zero of the gradient vector field) of a functional J. The straight forward generalization to functionals defined on a manifold is given below.

Definition 1.3.7. For a Riemannian manifold M and a two times differentiable functional $J\colon M \to \mathbb{R}$ the *Riemannian Newton method* is given by the iteration

$$\phi_R(p) := \exp_p\left(-(\operatorname{Hess} J(p))^{-1} \operatorname{grad} J(p)\right).$$

If $M \subset \mathbb{R}^n$ is a Riemannian submanifold of $\mathbb{R}^n$ the *projection-based Newton method* is given by the iteration

$$\phi_{\mathcal{P}}(p) := e_p\left(-(\operatorname{Hess} J(p))^{-1} \operatorname{grad} J(p)\right) = \mathcal{P}\left(p - (\operatorname{Hess} J(p))^{-1} \operatorname{grad} J(p)\right).$$

To prove local quadratic convergence we will need the following lemma. It can be regarded as a weak version of the fact that the derivative of the gradient is the Hessian.

Lemma 1.3.8. *Let M be a Riemannian manifold, p^* a critical point of a three times differentiable functional $J\colon M \to \mathbb{R}$ and $p \in M$. Then we have*

$$\left|\operatorname{grad} J(p) + \operatorname{Hess} J(p) log_p(p^*)\right| \lesssim d^2(p, p^*). \tag{1.10}$$

The implicit constant depends only on the third derivative of $J \circ exp_p$ at 0.

Proof. Let $v := \log_p(p^*)$ and $w := \operatorname{grad} J(p) + \operatorname{Hess} J(p)v$. Replacing v in (1.4) by $v + tw$ and using that $\operatorname{Hess} J(p)$ is self-adjoint yields

$$\begin{aligned}
J(\exp_p(v+tw)) &= J(p) + \langle v + tw, \operatorname{grad} J(p)\rangle + \frac{1}{2}\langle v + tw, \operatorname{Hess} J(p)(v+tw)\rangle + \mathcal{O}(|v+tw|^3) \\
&= J(p) + \langle v, \operatorname{grad} J(p)\rangle + \frac{1}{2}\langle v, \operatorname{Hess} J(p)v\rangle + \mathcal{O}(|v|^3) \\
&\quad + t\left(\langle w, \operatorname{grad} J(p) + \operatorname{Hess} J(p)v\rangle + \mathcal{O}(|w||v|^2)\right) + \mathcal{O}(t^2) \\
&= J(p^*) + t\left(|w|^2 + \mathcal{O}(|w||v|^2)\right) + \mathcal{O}(t^2).
\end{aligned}$$

Since p^* is critical point of J we have $J(\exp_p(v+tw)) = J(p^*) + \mathcal{O}(t^2)$ hence $|w|^2 = \mathcal{O}(|w||v|^2)$ and therefore $|w| \lesssim |v|^2$, which is equivalent to (1.10). □

We can now prove quadratic convergence.

Theorem 1.3.9. *Let M be a Riemannian manifold, p^* a critical point of a three times differentiable functional $J\colon M \to \mathbb{R}$, $p \in M$ and $\phi_R(p)$ as defined in Definition* (1.3.7)*. Assume that* $\operatorname{Hess} J(p^*)$ *defined in Section 1.3.1 is invertible. Then we have*

$$d(\phi_R(p), p^*) \lesssim d^2(p, p^*)$$

for all p in a neighborhood of p^*. The implicit constant depends only on the third derivative of $J \circ exp_{p^*}$ at 0 and on the operator norm of $(\operatorname{Hess} J(p^*))^{-1}$.

Proof. Let v be as in Lemma 1.3.8. By differentiability of J and Lemma 1.3.8 we have

$$\begin{aligned}
\left|(\operatorname{Hess} J(p))^{-1} \operatorname{grad} J(p) + v\right| &\lesssim \left\|(\operatorname{Hess} J(p))^{-1}\right\|_{T_pM \to T_pM} |\operatorname{grad} J(p) + \operatorname{Hess} J(p)v| \\
&\lesssim \left(\left\|(\operatorname{Hess} J(p^*))^{-1}\right\|_{T_pM \to T_pM} + \mathcal{O}(d(p, p^*))\right) d^2(p, p^*) \\
&\lesssim d^2(p, p^*).
\end{aligned}$$

We now have

$$\begin{aligned}
d\left(\phi(p), p^*\right) &= d\left(\exp_p(-(\operatorname{Hess} J(p))^{-1} \operatorname{grad} J(p)), \exp_p(v)\right) \\
&\lesssim \left|(\operatorname{Hess} J(p))^{-1} \operatorname{grad} J(p) + v\right| \\
&\lesssim d^2(p, p^*). \quad \square
\end{aligned}$$

Quadratic convergence of the projection-based Newton method can be proven analogously. Other proofs of quadratic convergence can be found in [3] and references therein.

Implementing the Riemannian Newton method

To evaluate the iteration of the Riemannian Newton method we need to compute the solution $v \in T_pM$ of the system $\operatorname{Hess} J(p)v = -\operatorname{grad} J(p)$. The following proposition shows how this can be done using representations $g \in \mathbb{R}^n$ and $H \in \mathbb{R}^{n \times n}$ of $\operatorname{grad} J$ and $\operatorname{Hess} J$.

Proposition 1.3.10. *Let M be a Riemannian submanifold of $\mathbb{R}^n$, $p \in M$, J a two times differentiable functional with* $\operatorname{Hess} J(p)$ *invertible, $g \in \mathbb{R}^n$ respectively $H \in \mathbb{R}^{n \times n}$ a representation of* $\operatorname{grad} J(p)$ *respectively* $\operatorname{Hess} J(p)$*, $S \in \mathbb{R}^{n \times \dim(M)}$ with*

$$\{Sx \mid x \in \mathbb{R}^{\dim(M)}\} = T_pM$$

and $x \in \mathbb{R}^{\dim(M)}$ the solution of the linear equation

$$S^T Sym(H) S x = -S^T g, \tag{1.11}$$

where $Sym(H) = (H + H^T)/2$ *is the symmetrization of* H*. Then* $u := Sx$ *is the solution of*

$$\operatorname{Hess} J(p)u = -\operatorname{grad} J(p). \tag{1.12}$$

Proof. Since $\operatorname{Hess} J(p)u \in T_pM$ and $\operatorname{grad} J(p) \in T_pM$ it is enough to prove that

$$\langle v, \operatorname{Hess} J(p)u\rangle = \langle v, -\operatorname{grad} J(p)\rangle \text{ for all } v \in T_pM.$$

Let $y \in \mathbb{R}^{\dim(M)}$ be such that $Sy = v$. By Proposition 1.3.4 we have

$$\langle v, \operatorname{Hess} J(p)u\rangle = \langle v, \operatorname{Sym}(H)u\rangle = \langle Sy, \operatorname{Sym}(H)Sx\rangle = y^T S^T \operatorname{Sym}(H)Sx$$

$$= -y^T S^T g = \langle Sy, -g\rangle = \langle v, -\operatorname{grad} J(p)\rangle. \quad \square$$

1.4 Averages of manifold-valued data

Assume we are given manifold-valued data $(p_i)_{i=1}^m \subset M$ where $m \in \mathbb{N}$. A key operation we want to perform with such data is to build weighted averages. We study two natural generalizations of the weighted average $\sum_{i=1}^m w_i p_i$, the Riemannian average in Section 1.4.1 and the projection average in Section 1.4.2.

1.4.1 Riemannian average

Note that for $(p_i)_{i=1}^m \subset \mathbb{R}^n$ and $(w_i)_{i=1}^m \subset \mathbb{R}$ with $\sum_{i=1}^m w_i = 1$ we have

$$\operatorname*{arg\,min}_{p\in\mathbb{R}^n} \sum_{i=1}^m w_i |p - p_i|^2 = \sum_{i=1}^m w_i p_i.$$

The idea of the Riemannian average, also known as the Karcher mean, is to replace $|p - p_i|$ with $d_g(p, p_i)$.

Definition 1.4.1. For manifold-valued data $(p_i)_{i=1}^m \subset M$ and weights $(w_i)_{i=1}^m \subset \mathbb{R}$ with $\sum_{i=1}^m w_i = 1$ the *Riemannian average* is defined by

$$av_{Riem}((p_i)_{i=1}^m, (w_i)_{i=1}^m) := \operatorname*{arg\,min}_{p\in M} J_R(p)$$

where

$$J_R(p) := \sum_{i=1}^m w_i d_g^2(p, p_i).$$

In [28] it is shown that:

- We have
$$\operatorname{grad}_p d_g^2(p,q) = -2\log_p(q). \tag{1.13}$$
and hence
$$\operatorname{grad}_p J_R(p) = -2\sum_{i=1}^{m} w_i \log(p,p_i). \tag{1.14}$$

- If M has nonpositive (sectional) curvature, or the points $(p_i)_{i=1}^m$ are close enough to each other, the gradient vector field (1.14) has a unique zero in a neighborhood of the points $(p_i)_{i=1}^m$. Hence, the Riemannian average is in these cases well-defined and can also be characterized by the condition $\operatorname{grad}_p J_R(p) = 0$. Furthermore, there are some estimates which emphasize the use of gradient descent with step size $\frac{1}{2}$ (see Section 1.3.2) to numerically compute the Riemannian average.

Gradient descent with step size $\frac{1}{2}$ for J_R is given by the iteration

$$\phi(p) := \exp_p\left(-\frac{1}{2}\operatorname{grad} J_R(p)\right) = \exp_p\left(\sum_{i=1}^{m} w_i \log_p(p_i)\right). \tag{1.15}$$

In the next proposition, we make a statement about the convergence rate of this iteration. This statement explains why only a few iterations are necessary to compute the Riemannian average up to high precision. A more detailed analysis has been performed in [54].

Proposition 1.4.2. *Let M be a Riemannian manifold, $p \in M$, $(w_i)_{i=1}^m \subset \mathbb{R}$ with sum 1 and $h > 0$ be small enough such that the Riemannian average $p^* = av_{Riem}((p_i)_{i=1}^m, (w_i)_{i=1}^m)$ is well-defined for any $(p_i)_{i=1}^m \subset B_h(p) := \{q \in M \mid d_g(p,q) < h\}$. Then we have*

$$d_g(\phi(q), p^*) \lesssim h^2 d_g(q, p^*)$$

for all $q \in B_h(p)$ where the implicit constant depends only on the (sectional) curvature of M.

Proof. By (1.5.1)-(1.5.3) of [28] we have

$$d_g(\phi(q),p^*) \leq \frac{1}{1-C_1h^2}|\operatorname{grad} J(\phi(q))| \leq \frac{C_2h^2}{1-C_1h^2}|\operatorname{grad} J(q)| \leq \frac{C_2h^2(1+C_3h^2)}{1-C_1h^2} d_g(q,p^*),$$

where C_1, C_2, C_3 are constants depending only on the (sectional) curvature of M. □

1.4.2 Projection average

If $M \subset \mathbb{R}^n$ is a Riemannian submanifold we can define an average which in many cases is easier to compute

Definition 1.4.3. For $(p_i)_{i=1}^m \subset M$ where M is a Riemannian submanifold of $\mathbb{R}^n$ and weights $(w_i)_{i=1}^m \subset \mathbb{R}$ with $\sum_{i=1}^m w_i = 1$ the *projection average* is defined by

$$av_{\mathcal{P}}((p_i)_{i=1}^m, (w_i)_{i=1}^m) := \mathcal{P}\left(\sum_{i=1}^m w_i p_i\right).$$

The next proposition shows that the projection average could have also be defined in a similar way as the Riemannian average in Definition 1.4.1. We only have to replace the geodesic distance with the Euclidean distance.

Proposition 1.4.4. *We have*

$$\arg\min_{p\in M} J_{\mathcal{P}}(p) = av_{\mathcal{P}}((p_i)_{i=1}^m, (w_i)_{i=1}^m),$$

where

$$J_{\mathcal{P}}(p) := \sum_{i=1}^m w_i |p - p_i|^2.$$

Proof. We have

$$\begin{aligned}
\arg\min_{p\in M} J_{\mathcal{P}}(p) &= \arg\min_{p\in M} \sum_{i=1}^m w_i |p-p_i|^2 \\
&= \arg\min_{p\in M} \sum_{i=1}^m w_i |p|^2 - 2\langle p, w_i p_i\rangle + w_i |p_i|^2 \\
&= \arg\min_{p\in M} |p|^2 - 2\left\langle p, \sum_{i=1}^m w_i p_i \right\rangle + \sum_{i=1}^m w_i |p_i|^2 \\
&= \arg\min_{p\in M} \left|p - \sum_{i=1}^m w_i p_i\right|^2 - \left|\sum_{i=1}^m w_i p_i\right|^2 + \sum_{i=1}^m w_i |p_i|^2 \\
&= \arg\min_{p\in M} \left|p - \sum_{i=1}^m w_i p_i\right| \\
&= \mathcal{P}\left(\sum_{i=1}^m w_i p_i\right) \\
&= av_{\mathcal{P}}((p_i)_{i=1}^m, (w_i)_{i=1}^m). \quad \square
\end{aligned}$$

1.4.3 Proximity of the Riemannian and the projection average

In this section, we show that the Riemannian and the projection average are numerically close to each other. More precisely, we show that if the data $(p_i)_{i=1}^m$ is contained in a ball of radius $h \ll 1$ then the error between the Riemannian and the projection average is of size h^3. This will be important later in the proof of Proposition 3.4.4 where we estimate the L^p-norm of the difference between the Riemannian average based approximation and the projection-based approximation of a function.

Proposition 1.4.5. *Let M be a Riemannian submanifold of $\mathbb{R}^n$, $p \in M$ and $(w_i)_{i=1}^m \subset \mathbb{R}$ with $\sum_{i=1}^m w_i = 1$. Assume that the closest point projection $\mathcal{P}$ is three times differentiable in a uniform neighborhood of M. Then if $h > 0$ is small enough we have for $(p_i)_{i=1}^m \subset B_h(p) \subset M$ that*

$$d_g\left(av_{Riem}((p_i)_{i=1}^m, (w_i)_{i=1}^m), av_{\mathcal{P}}((p_i)_{i=1}^m, (w_i)_{i=1}^m)\right) \lesssim h^3,$$

where the implicit constant depends only on $|\mathcal{P}|_{C^2}$ and $|\mathcal{P}|_{C^3}$.

Proof. Let $p_R := av_{Riem}((p_i)_{i=1}^m, (w_i)_{i=1}^m)$ and $p_{\mathcal{P}} := av_{\mathcal{P}}((p_i)_{i=1}^m, (w_i)_{i=1}^m)$. By (1.5.1) of [28] and Lemma A.2 we have that

$$d_g(p_R, p_{\mathcal{P}}) \leq (1+\mathcal{O}(h^2))|\operatorname{grad} J_R(p_{\mathcal{P}})| = (1+\mathcal{O}(h^2))\left(|\operatorname{grad} J_{\mathcal{P}}(p_{\mathcal{P}})| + \mathcal{O}(h^3)\right) \lesssim h^3. \quad \square$$

An interesting open problem is if a similar estimate also holds for the derivatives of the averages with respect to the weights $(w_i)_{i=1}^m$. This would allow us to estimate the $W^{l,p}$-norms for $l > 0$ of the difference between the Riemannian average based approximation and the projection-based approximation of a function. For the average of only two points we can prove that the first and second derivative can also be bounded by h^3.

Proposition 1.4.6. *Let M be a Riemannian submanifold of $\mathbb{R}^n$. Assume that the closest point projection $\mathcal{P}$ is three times differentiable in a uniform neighborhood of M. Then for $p_1, p_2 \in M$ with $|p_1 - p_2|$ small enough, $i \in \{0, 1, 2\}$ and $t \in [0, 1]$ we have*

$$\left|\frac{d^i}{dt^i}\left(av_{Riem}((p_1, p_2), (t, 1-t)) - av_{\mathcal{P}}((p_1, p_2), (t, 1-t))\right)\right| \lesssim |p_2 - p_1|^3,$$

where the implicit constant depends only on $|\mathcal{P}|_{C^2}$ and $|\mathcal{P}|_{C^3}$.

Proof. Consider the curves $\gamma_g, \gamma_p \colon [0, 1] \to M$ defined by

$$\gamma_g(t) := av_{Riem}((p_1, p_2), (1-t, t))$$

and $\gamma_p(t) := av_{\mathcal{P}}((p_1, p_2), (1-t, t)) = \mathcal{P}((1-t)p_1 + tp_2)$. Note that γ_g is a geodesic and therefore satisfies the geodesic equation (see (1.3))

$$\ddot{\gamma}_g(t) = \mathcal{P}''(\gamma_g(t))[\dot{\gamma}_g(t), \dot{\gamma}_g(t)]. \tag{1.16}$$

Differentiating γ_p twice yields

$$\ddot{\gamma}_p(t) = \mathcal{P}''((1-t)p_1 + tp_2)[p_2 - p_1, p_2 - p_1]. \tag{1.17}$$

From Lemma 1.2.9 we have

$$|\dot{\gamma}_g(t) - (p_2 - p_1)| \lesssim |p_2 - p_1|^2. \tag{1.18}$$

Hence we notice that both curves satisfy a similar differential equation. Let

$$\delta(t) \;:=\; \gamma_g(t) \;-\; \gamma_p(t).$$

Note that we need to prove that $\max_{t\in(0,1)} |\dot{\delta}(t)| \lesssim |p_2 - p_1|^3$ and $\max_{t\in(0,1)} |\ddot{\delta}(t)| \lesssim |p_2 - p_1|^3$. Taking the difference of (1.16) and (1.17) and using (1.18) yields

$$\begin{aligned}
|\ddot{\delta}(t)| \;&\leq\; |\mathcal{P}''(\gamma_g(t))[\dot{\gamma_g}(t) - (p_2 - p_1), \dot{\gamma_g}(t)]| \\
&\quad + |\mathcal{P}''(\gamma_g(t))[p_2 - p_1, \dot{\gamma_g}(t) - (p_2 - p_1)]| \\
&\quad + |(\mathcal{P}''(\gamma_g(t)) - \mathcal{P}''((1-t)p_1 + tp_2))\,[p_2 - p_1, p_2 - p_1]| \\
&\lesssim\; |p_2 - p_1|^3 + |\gamma_g(t) - ((1-t)p_1 + tp_2)||p_2 - p_1|^2. \\
&\lesssim\; |p_2 - p_1|^3.
\end{aligned}$$

As $\int_0^1 \dot{\delta}(t)dt = \delta(1) - \delta(0) = 0 - 0 = 0$ we have

$$0 = \int_0^1 \dot{\delta}(t)dt = \dot{\delta}(0) + \int_0^1 (1-t)\ddot{\delta}(t)dt,$$

and hence

$$|\dot{\delta}(0)| = \left| \int_0^1 (1-t)\ddot{\delta}(t)dt \right| \lesssim |p_2 - p_1|^3.$$

Therefore

$$|\dot{\delta}(t)| = \left| \dot{\delta}(0) + \int_0^t \ddot{\delta}(s)ds \right| \lesssim |p_2 - p_1|^3. \quad \square$$

1.5 Example manifolds

In this section, we consider several different manifolds M. Namely, the sphere S^n, the special orthogonal group $SO(n)$ and the set of positive definite matrices $SPD(n)$. These are the most common examples of manifolds we encounter currently in applications with manifold-valued data. We identify the tangent space at a point $p \in M$, geodesics, the logarithm and the exponential map, the (geodesic) distance function and the closest point projection (if the manifold is a Riemannian submanifold of $\mathbb{R}^n$) of these manifolds. We also discuss how to compute the gradient and the Hessian of the squared distance function. To do so there are basically two techniques. One is to use an extension of the squared distance function and then Proposition 1.3.5. The second technique is to use an extension of the logarithm and then Proposition 1.3.2. Depending on the manifold and the distance function one of these techniques is usually easier to perform.

1.5.1 The sphere

Together with the standard inner product the sphere $S^{n-1} := \{x \in \mathbb{R}^n \mid |x| = 1\}$ is a Riemannian manifold. It acts as a model manifold since it is one of simplest computationally tractable manifolds. However, there are also many applications with sphere-valued data, e.g. the chromaticity part of an RGB-image or liquid crystals.

Tangent space, geodesics, exponential and logarithm map on the sphere

The tangent space at a point $p \in S^{n-1}$ is $T_pS^n = \{q \in \mathbb{R}^n | p^T q = 0\}$. The geodesics are the unit speed parametrizations of the great circles. The exponential and logarithm map on the sphere are explicitly given by

$$\exp_p(v) = \cos(|v|)p + \frac{\sin(|v|)}{|v|}v \quad \text{and} \quad \log_p(q) = \frac{\arccos(\langle p, q\rangle)}{\sqrt{1-\langle p, q\rangle^2}}\left(q - \langle p, q\rangle p\right).$$

The closest point projection and the Taylor expansion of $\exp_p$ and e_p on the sphere

The closest point projection $\mathcal{P}$ is simply the normalization, i.e.

$$\mathcal{P}(x) := \frac{x}{|x|} \quad \text{for all } x \in \mathbb{R}^n \backslash \{0\}.$$

The Taylor expansion of the exponential map and projection-based retraction on the sphere are

$$\exp_p(v) = \cos(|v|)p + \frac{\sin(|v|)}{|v|}v = p + v - \frac{|v|^2}{2}p + o(|v|^2) \text{ and} \tag{1.19}$$

$$e_p(v) = \mathcal{P}(p+v) = \frac{p+v}{|p+v|} = \frac{p+v}{|p| + \frac{|v|^2}{2|p|} + \mathcal{O}(|v|^4)} = p + v - \frac{|v|^2}{2}p + o(|v|^2).$$

It is by Proposition 1.2.8 no surprise that the Taylor expansions of exp and e_p coincide. By (1.7) we have

$$\mathcal{P}''(p)[v, v] = -|v|^2 p \text{ for all } v \in T_pM.$$

Hessian of a functional on the sphere

For a functional $J\colon (S^{n-1})^N \to \mathbb{R}$ and an extension $\bar{J}\colon U \subset (\mathbb{R}^n)^N \to \mathbb{R}$ we have by Proposition 1.3.5

$$\begin{aligned}\langle v, \operatorname{Hess} J(p)v\rangle &= \langle v, \operatorname{Hess} \bar{J}(p)v\rangle + \sum_{i=1}^N \langle \operatorname{grad}_i \bar{J}(p), \mathcal{P}''(p_i)[v_i, v_i]\rangle \\ &= \langle v, \operatorname{Hess} \bar{J}(p)v\rangle - \sum_{i=1}^N \langle \operatorname{grad}_i \bar{J}(p), p_i\rangle |v_i|^2.\end{aligned}$$

The Hessian $\operatorname{Hess} J(p)$ of J can therefore be represented by the symmetric matrix

$$\operatorname{Hess} \bar{J}(p) - \begin{pmatrix} \langle \operatorname{grad}_1 \bar{J}(p), p_1\rangle & & 0 \\ & \ddots & \\ 0 & & \langle \operatorname{grad}_N \bar{J}(p), p_N\rangle \end{pmatrix} \otimes I_n \tag{1.20}$$

where I_n is the $n \times n$ identity matrix and $\otimes$ denotes the Kronecker product.

Gradient and Hessian of the squared distance function on the sphere

We will now compute the gradient and the Hessian of functionals $J(p,q) := \alpha(p^T q)$ with $\alpha \in C^2((-1,1])$. Note that the squared geodesic respectively the squared Euclidean distance are of this form with $\alpha(x) = \arccos^2(x)$ respectively $\alpha(x) = 2 - 2x$.

Let $\bar{J}\colon (\mathbb{R}^{n+1})^2 \to \mathbb{R}$ be the natural extension of J defined by $\bar{J}(p,q) := \alpha(p^T q)$. The gradient of $\bar{J}$ is

$$\operatorname{grad} \bar{J}(p,q) = (\operatorname{grad}_p \bar{J}, \operatorname{grad}_q \bar{J}) = \alpha'(p^T q)(q,p).$$

The Hessian $\operatorname{Hess} J(p)$ can be represented by the symmetric matrix in (1.20) which boils down to

$$\alpha''(p^T q) \begin{pmatrix} q \\ p \end{pmatrix} \begin{pmatrix} q^T & p^T \end{pmatrix} + \begin{pmatrix} -\beta(p^T q) & \alpha'(p^T q) \\ \alpha'(p^T q) & -\beta(p^T q) \end{pmatrix} \otimes I_{n+1},$$

where $\beta(t) := t\alpha'(t)$.

1.5.2 The special orthogonal group $SO(n)$

The set of $n \times n$ orthogonal matrices with determinant 1 together with the inner product

$$\langle A, B \rangle := \operatorname{tr}(A^T B) = \sum_{i,j=1}^{n} a_{ij} b_{ij} \text{ for all } A = (a_{ij})_{i,j=1}^n, B = (b_{ij})_{i,j=1}^n \subset \mathbb{R}^{n \times n}, \quad (1.21)$$

where tr denotes the trace, is a Riemannian manifold of dimension $n\ (n\ -\ 1)\ /\ 2$. If we identify $\mathbb{R}^{n \times n}$ with $\mathbb{R}^{n^2}$ the inner product becomes simply the standard inner product. Hence we can see $SO(n)$ as a Riemannian submanifold of $\mathbb{R}^{n^2}$. Data with values in $SO(n)$ occur for example in Cosserat-type material models [36, 46, 40, 37]. Note that the dimension of the manifold is $n(n-1)/2$ which is significantly lower than n^2, the dimension of the space it is embedded in. Luckily, the dimension of the system of equations in optimization methods for manifold valued data (cf. (1.11)) depends only on the dimension of the manifold and not on the dimension of the space it is embedded in.

Tangent space, geodesics, exponential and logarithm map on $SO(n)$

The tangent space at $A \in SO(n)$ is

$$T_A SO(n) = \{AS | S \in \mathbb{R}^{n \times n}\ S^T = -S\}.$$

The exponential and logarithm map are given by

$$\exp_A(X) = A\operatorname{Exp}(A^T X) \quad \text{and} \quad \log_A(B) = A\operatorname{Log}(A^T B)$$

where Exp respectively Log denotes the matrix exponential respectively matrix logarithm. The geodesic distance function is given by $d_g(A,B) = \|\log_A(B)\|_F = \|\mathrm{Log}(A^TB)\|_F$, where the subscript F denotes the Frobenius norm defined by

$$\|A\|_F := \sqrt{\mathrm{tr}(A^TA)} = \sqrt{\sum_{i,j=1}^{n} a_{ij}^2}.$$

Gradient and Hessian of the squared distance function on $SO(n)$

We will now explain how to compute the gradient and the Hessian of the squared distance function $J\colon (SO(n))^2 \to \mathbb{R}$ defined by $J(A,B) := d_g^2(A,B)$ for all $A,B \in SO(n)$. By (1.13) the gradient is given by

$$\mathrm{grad}\, J(A,B) = (-2\log_A(B), -2\log_B(A)) = (-2A\mathrm{Log}(A^TB), -2B\mathrm{Log}(B^TA)).$$

By Proposition 1.3.2 we can compute the classical derivatives of this expression to get a representation for the Hessian. To approximately compute it one can use the chain and product rule. To compute the derivative of the matrix logarithm one can use the matrix identities

$$\mathrm{Log}(X^2) = 2\mathrm{Log}(X) \quad \text{and} \quad \mathrm{Log}(X) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}(X - I_n)^k \tag{1.22}$$

that holds for all $X \in \mathbb{R}^{n\times n}$ (the sum converges whenever the spectral radius of $X - I_n$ is smaller than 1) as well as

$$\frac{\partial (X^k)_{ij}}{\partial X_{mo}} = \sum_{l=0}^{k-1} (X^l)_{im}(X^{k-l-1})_{oj}$$

that holds for all $X \in \mathbb{R}^{n\times n}$ and $i,j,m,o \in \{1,\dots,n\}$ for the derivative of the matrix power. The matrix exponential (respectively matrix logarithm) can be computed by diagonalizing the matrix and applying the exponential (respectively logarithm) to the eigenvalues.

1.5.3 The compact Stiefel manifold

For $k < n$ the orthogonal Stiefel manifold is defined by

$$St(k,n) = \{A \in \mathbb{R}^{n\times k} | A^TA = I_k\},$$

where I_k denotes the $k \times k$ identity matrix. Together with the inner product (1.21) this is a Riemannian manifold. As special cases the Stiefel manifold contains the sphere ($k = 1$) and the special orthogonal group ($k = n$).

The closest point projection and its second derivative on $St(k,n)$

The closest point projection of a matrix $A \in \mathbb{R}^{n\times k}$ onto $St(k,n)$ is given by dropping the diagonal matrix Σ of the reduced singular value decomposition $A = U\Sigma V^T$, i.e. $\mathcal{P}(A) = UV^T$. We can also write the projection in the explicit form $\mathcal{P}(A) = A(A^TA)^{-\frac{1}{2}}$. To numerically compute $\mathcal{P}(A)$ without computing eigenvalues and eigenvectors one can use the iteration $\phi\colon \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times k}$ defined by

$$\phi(X) := \frac{1}{2}X\left(I_k + (X^TX)^{-1}\right).$$

This iteration is based on Heron's method for the computation of $\sqrt{1}$ and converges quadratically [25] to $\mathcal{P}(A)$. Note that the tangent space $T_ASt(k,n)$ is given by

$$T_ASt(k,n) = \{X \in R^{n\times k} | A^TX + X^TA = 0\}.$$

For $A \in St(k,n)$ and $X \in T_ASt(k,n)$ we therefore get

$$\begin{aligned}
\mathcal{P}(A+X) &= (A+X)\left((A+X)^T(A+X)\right)^{-\frac{1}{2}} \\
&= (A+X)\left(I_k + X^TX\right)^{-\frac{1}{2}} \\
&= (A+X)(I_k - \frac{1}{2}X^TX + \mathcal{O}(|X|^4)) \\
&= A + X - \frac{1}{2}AX^TX + \mathcal{O}(|X|^3)
\end{aligned}$$

Hence we have

$$\mathcal{P}''(A)[X,X] = -AX^TX.$$

The Riemannian Hessian on $St(k,n)$

For a functional $J\colon (St(k,n))^N \to \mathbb{R}$ and an extension $\bar{J}\colon U \subset (\mathbb{R}^{n\times k})^N \to \mathbb{R}$ we have by Proposition 1.3.5 for $A = (A_i)_{i=1}^N \in (St(k,n))^N$ and $X = (X_i)_{i=1}^N \in T_A(St(k,n))^N$ that

$$\begin{aligned}
\langle X, \operatorname{Hess} J(A)X\rangle &= \langle X, \operatorname{Hess} \bar{J}(A)X\rangle + \sum_{i=1}^N \langle \operatorname{grad}_i \bar{J}(A), \mathcal{P}''(A_i)[X_i,X_i]\rangle \\
&= \langle X, \operatorname{Hess} \bar{J}(A)X\rangle - \sum_{i=1}^N \langle \operatorname{grad}_i \bar{J}(A), A_iX_i^TX_i\rangle \\
&= \langle X, \operatorname{Hess} \bar{J}(A)X\rangle - \sum_{i=1}^N \langle A_i^T \operatorname{grad}_i \bar{J}(A), X_i^TX_i\rangle.
\end{aligned}$$

The Hessian $\operatorname{Hess} J(A)$ of J at $A \in (St(k,n))^N$ can therefore be represented by the matrix

$$H := \operatorname{Hess} \bar{J}(A) - \begin{pmatrix} A_1^T \operatorname{grad}_1 \bar{J}(A) & & 0 \\ & \ddots & \\ 0 & & A_N^T \operatorname{grad}_N \bar{J}(A) \end{pmatrix} \otimes I_n, \tag{1.23}$$

where $\otimes$ denotes the Kronecker product. Note that (1.23) is a generalization of (1.20). The matrix H of (1.23) is in general not symmetric. As explained in Section 1.3.3 to apply the Newton method we will use the symmetrization of H. In numerical experiments it was however observed that the symmetrization is actually not necessary. Even though (1.12) is then no longer true, the corresponding iteration seems to converge quadratically to the minimizer of J.

1.5.4 The space of positive definite matrices $SPD(n)$

The space of $n \times n$ positive definite matrices $SPD(n)$ is a manifold. Data with values in $SPD(n)$ occur for example in diffusion tensor magnetic resonance imaging (DT-MRI). For $A \in SPD(n)$, the tangent space $T_A SPD(n)$ at A can be identified by the space of symmetric matrices.

Why the standard inner product is not a good choice

There are different choices for the inner product on the tangent space each resulting in a different Riemannian manifold. One choice would be to take the same inner product as for the special orthogonal group (1.21). Since $SPD(n)$ is an open subset of the Euclidean space of symmetric matrices this would lead to very simple computations. However, for many applications this inner product is not a good choice. The determinant of an average of matrices can then be much larger than the determinant of the original matrices. Consider for example $0 < \epsilon \ll 1$ and the two SPD-matrices $A = diag(1, \epsilon)$ and $B = diag(\epsilon, 1)$, which have determinant ϵ. The average with respect to the metric induced by the inner product (1.21) would be $(A + B)/2 = diag((1 + \epsilon)/2, (1 + \epsilon)/2)$ and has a significantly higher determinant, close to $1/4$. In DT-MRI the determinant is related to the amount of diffusion and hence averaging would introduce more diffusion which is physically unrealistic. This is also known as the swelling effect.

A better choice for the inner product, geodesics, exponential and logarithm map

Due to the problems of the preceding section a different inner product given by

$$\langle X, Y \rangle_A := \operatorname{tr}(A^{-1} X A^{-1} Y), \text{ for all } X, Y \in T_A SPD(n), \tag{1.24}$$

is usually considered. A detailed analysis of the geometry of the resulting space can be found in [35]. The inner product (1.24) induces the metric $d_g \colon SPD(n) \times SPD(n) \to \mathbb{R}_{\geq 0}$ given by

$$d_g(A, B) = \|\mathrm{Log}(A^{-1/2} B A^{-1/2})\|_F. \tag{1.25}$$

By [35] the exponential map on the space of SPD-matrices is given by

$$\exp_A(X) = A^{\frac{1}{2}} \mathrm{Exp}(A^{-\frac{1}{2}} X A^{-\frac{1}{2}}) A^{\frac{1}{2}}.$$

Hence, the logarithm map is given by

$$\log_A(B) = A^{\frac{1}{2}} \mathrm{Log}(A^{-\frac{1}{2}} B A^{-\frac{1}{2}}) A^{\frac{1}{2}}.$$

The geodesic connecting A and B is therefore given by

$$\gamma(t) = \exp_A(t \log_A(B)) = A^{\frac{1}{2}} \mathrm{Exp}(t \mathrm{Log}(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})) A^{\frac{1}{2}}.$$

Because of the identity $\det(\mathrm{Exp}(A)) = e^{\mathrm{tr}(A)}$ we have

$$\det(\gamma(t)) = \det(A)^{1-t} \det(B)^t.$$

Hence, the problem of the inner product of Section 1.5.4 does not occur for the inner product (1.24).

Hessian of the squared distance function on $SPD(n)$

We will now explain how to compute the gradient and Hessian of the squared distance function $J \colon (SPD(n))^2 \to \mathbb{R}$ defined by $J(A, B) := d_g^2(A, B)$ for all $A, B \in SPD(n)$. By (1.13) the gradient with respect to the inner product (1.24) is given by

$$\begin{aligned} \mathrm{grad}\, J(A, B) &= (-2\log_A(B), -2\log_B(A)) \\ &= (-2A^{\frac{1}{2}} \mathrm{Log}(A^{-\frac{1}{2}} B A^{-\frac{1}{2}}) A^{\frac{1}{2}}, -2B^{\frac{1}{2}} \mathrm{Log}(B^{-\frac{1}{2}} A B^{-\frac{1}{2}}) B^{\frac{1}{2}}). \end{aligned}$$

For all $Z \in T_A SPD(n)$ we have

$$\begin{aligned} \langle \log_A(B), Z \rangle_A &= \langle A^{\frac{1}{2}} \mathrm{Log}(A^{-\frac{1}{2}} B A^{-\frac{1}{2}}) A^{\frac{1}{2}}, Z \rangle_A \\ &= \langle A^{-1} A^{\frac{1}{2}} \mathrm{Log}(A^{-\frac{1}{2}} B A^{-\frac{1}{2}}) A^{\frac{1}{2}} A^{-1}, Z \rangle \\ &= \langle A^{-\frac{1}{2}} \mathrm{Log}(A^{-\frac{1}{2}} B A^{-\frac{1}{2}}) A^{-\frac{1}{2}}, Z \rangle. \end{aligned}$$

The classical gradient with respect to the Euclidean distance is therefore

$$A^{-\frac{1}{2}} \mathrm{Log}(A^{-\frac{1}{2}} B A^{-\frac{1}{2}}) A^{-\frac{1}{2}}.$$

To compute the classical Hessian we simply compute the classical derivative of this expression. As explained in Section 1.5.2 one can use the chain rule, product rule as well as the identities (1.22) to compute the derivative. To compute the derivative of the

square root of a matrix observe that $Y^{k,l} := d\sqrt{X}/dX_{kl} \in \mathbb{R}^{n\times n}$ is the solution of the Lyapunov equation

$$Y^{k,l}\sqrt{X} + \sqrt{X}Y^{k,l} = E_{kl}, \tag{1.26}$$

where $E_{kl} \in \mathbb{R}^{n\times n}$ is the matrix defined by $(E_{kl})_{mo} := \delta_{km}\delta_{lo}$. The derivative of the matrix inverse is given by

$$\frac{d(X^{-1})_{ij}}{dX_{kl}} = -(X^{-1})_{ik}(X^{-1})_{lj}.$$

Since we compute the classical derivatives we will also use the classical Newton method to optimize J. However, since $(A,B) - \operatorname{Hess} J(A,B)^{-1} \operatorname{grad} J(A,B)$ can have non SPD-matrices as values it is recommended to use the iteration

$$(A,B) \mapsto \exp_{(A,B)}(-\operatorname{Hess} J(A,B)^{-1} \operatorname{grad} J(A,B))$$

instead. From $\exp_A(X) = A^{\frac{1}{2}}\operatorname{Exp}(A^{-\frac{1}{2}}XA^{-\frac{1}{2}})A^{\frac{1}{2}} = A + X + \mathcal{O}(|X|^2)$ it follows that the method still has quadratic convergence.

The Log-Euclidean distance

Since the inner product (1.24) and the induced metric (1.25) are rather cumbersome to use, it can be worthwhile to use an alternative metric which in practice yields a similar or comparable result but is easier to work with. Note that the matrix logarithm Log defines a bijection between the set of positive definite matrices and the set of symmetric matrices. A common idea to solve problems where positive definite matrices are involved is to first map all data to the space of symmetric matrices by the matrix logarithm Log, then solve the problem in the Euclidean space of symmetric matrices and finally map the data back by the matrix exponential Exp. This procedure is equivalent with using the Log-Euclidean distance

$$d_{LE}(A,B) := \|\operatorname{Log}(A) - \operatorname{Log}(B)\|_F.$$

A crucial question is how far apart the Log-Euclidean distance is from the metric induced by the inner product (1.24). Note that if two matrices A and B commute we have $d_{LE}(A,B) = d_g(A,B)$. The following proposition shows that the Log-Euclidean distance can be bounded by the geodesic metric.

Proposition 1.5.1. *For any $A, B \in SPD(n)$ we have*

$$d_{LE}(A,B) \leq d_g(A,B).$$

Proof. By [52] the eigenvalues of $\operatorname{Exp}(R+S)$ are majorized by the eigenvalues of

$$\operatorname{Exp}(R/2)\operatorname{Exp}(S)\operatorname{Exp}(R/2)$$

for all symmetric matrices $R, S \in \mathbb{R}^{n\times n}$ in a logarithmic sense, i.e. if $\lambda_1 \geq \cdots \geq \lambda_n > 0$ are the eigenvalues of $\mathrm{Exp}(R+S)$ and $\mu_1 \geq \mu_2 \cdots \geq \mu_n > 0$ are the eigenvalues of $\mathrm{Exp}(R/2)\mathrm{Exp}(S)\mathrm{Exp}(R/2)$ we have

$$\sum_{i=1}^{k} \log(\lambda_i) \leq \sum_{i=1}^{k} \log(\mu_i) \text{ for all } 1 \leq k \leq n-1$$

and $\sum_{i=1}^{n} \log(\lambda_i) = \sum_{i=1}^{n} \log(\mu_i)$. It follows that

$$\|R+S\|_F^2 = \sum_{i=1}^{n} (\log(\lambda_i))^2 \leq \sum_{i=1}^{n} (\log(\mu_i))^2 = \|\mathrm{Log}(\mathrm{Exp}(R/2)\mathrm{Exp}(S)\mathrm{Exp}(R/2))\|_F^2.$$

Choosing $R = -\mathrm{Log}(A)$ and $S = \mathrm{Log}(B)$ yields

$$d_{LE}(A,B) = \|\mathrm{Log}(B) - \mathrm{Log}(A)\|_F \leq \|\mathrm{Log}(A^{-1/2}BA^{-1/2})\|_F = d_g(A,B). \quad \square$$

By the following proposition the two metrics are however not strongly equivalent.

Proposition 1.5.2. *For $n \geq 2$ and $C > 0$ there exists positive definite matrices $A, B \in SPD(n)$ such that*

$$d_g(A,B) > C d_{LE}(A,B).$$

Proof. Consider the family of matrices

$$A(t) := \mathrm{Exp}\begin{pmatrix} t & 0 \\ 0 & 0 \end{pmatrix} \text{ and } B(t) := \mathrm{Exp}\begin{pmatrix} t & 1 \\ 1 & 0 \end{pmatrix}.$$

Then we have $d_{LE}(A(t), B(t)) = \sqrt{2}$ for all $t \in \mathbb{R}$. To show our claim we prove that $d_g(A(t), B(t))$ gets arbitrarily large for $t \to \infty$. The eigenvalues of $\begin{pmatrix} t & 1 \\ 1 & 0 \end{pmatrix}$ are $\lambda_1 = t + t^{-1} + \mathcal{O}(t^{-3})$ and $\lambda_2 = -t^{-1} + \mathcal{O}(t^{-3})$. Corresponding eigenvectors are $(\lambda_1, 1) = (t + t^{-1}, 1) + \mathcal{O}(t^{-3})$ and $(-1, t - \lambda_2) = (-1, t + t^{-1}) + \mathcal{O}(t^{-3})$. We get the approximate diagonalization

$$\begin{aligned} \begin{pmatrix} t & 1 \\ 1 & 0 \end{pmatrix} &= S \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} S^{-1} + \mathcal{O}(t^{-3}) \\ &= \begin{pmatrix} t+t^{-1} & -1 \\ 1 & t+t^{-1} \end{pmatrix} \begin{pmatrix} t+t^{-1} & 0 \\ 0 & -t^{-1} \end{pmatrix} \frac{1}{t^2+3+t^{-2}} \begin{pmatrix} t+t^{-1} & 1 \\ -1 & t+t^{-1} \end{pmatrix} \\ &\quad + \mathcal{O}(t^{-3}). \end{aligned}$$

For the matrix exponential we have

$$B(t) = \mathrm{Exp}\begin{pmatrix} t & 1 \\ 1 & 0 \end{pmatrix}$$

$$
\begin{aligned}
&= \begin{pmatrix} t+t^{-1} & -1 \\ 1 & t+t^{-1} \end{pmatrix} \begin{pmatrix} e^t\left(1+t^{-1}+\frac{1}{2}t^{-2}\right) & 0 \\ 0 & 1 \end{pmatrix} \frac{1}{t^2+3+t^{-2}} \begin{pmatrix} t+t^{-1} & 1 \\ -1 & t+t^{-1} \end{pmatrix} \\
&\quad +\mathcal{O}(t^{-3}e^t) \\
&= \begin{pmatrix} 1 & t^{-1} \\ t^{-1} & t^{-2} \end{pmatrix} e^t + e^t \begin{pmatrix} \mathcal{O}(t^{-1}) & \mathcal{O}(t^{-2}) \\ \mathcal{O}(t^{-2}) & \mathcal{O}(t^{-3}) \end{pmatrix}.
\end{aligned}
$$

Therefore

$$
A(t)^{-1/2}B(t)A(t)^{-1/2} = \begin{pmatrix} 1 & t^{-1}e^{\frac{t}{2}} \\ t^{-1}e^{\frac{t}{2}} & t^{-2}e^t \end{pmatrix} + \begin{pmatrix} \mathcal{O}(t^{-1}) & \mathcal{O}(t^{-2}e^{\frac{t}{2}}) \\ \mathcal{O}(t^{-2}e^{\frac{t}{2}}) & \mathcal{O}(t^{-3}e^t) \end{pmatrix}.
$$

It follows that the entries of $A(t)^{-1/2}B(t)A(t)^{-1/2}$ get arbitrarily large and therefore also the Frobenius norm of its logarithm. □

2 Total Variation Minimization

In this chapter, we consider the problem of restoring a noisy image $u^n \colon V_k \subset V \to M$ where V is the set of pixels and V_k is a subset where our noisy image is defined. To perform this task, we consider the total variation[1] (TV) of an image $u \colon V \to M$ defined by

$$TV(u) := \sum_{(i,j)\in E} d(u_i, u_j), \tag{2.1}$$

where $d \colon M \times M \to \mathbb{R}_{\geq 0}$ is a metric on M, e.g. the geodesic or the Euclidean distance. If $V = \{1, \dots, n\}$ and $E = \{(i, i+1) \mid i \in \{1, \dots, n-1\}\}$ our total variation (2.1) coincides for example with the well-known discrete one-dimensional TV. A key observation is that the TV of a typical image is relatively small, while adding noise increases the TV significantly. To restore the image, we seek an image $u \in M^V$ (i.e. $u \colon V \to M$) which has relatively small TV while still being "close" to the noisy image $u^n \in M^{V_k}$. To achieve this, we minimize the TV functional $J \colon M^V \to \mathbb{R}$ defined by

$$J(u) := \frac{1}{2} \sum_{i\in V_k} d^2(u_i, u_i^n) + \lambda TV(u) = \frac{1}{2} \sum_{i\in V_k} d^2(u_i, u_i^n) + \lambda \sum_{(i,j)\in E} d(u_i, u_j), \tag{2.2}$$

where $\lambda > 0$ is a positive constant balancing the two parts of the functional. The TV part of J is not differentiable because the geodesic distance $d \colon M \times M \to \mathbb{R}_{\geq 0}$ is not differentiable on the diagonal. Therefore, standard methods to minimize J (e.g. Newton) will fail. Thus, for $\epsilon > 0$ we define a regularized functional $J^\epsilon \colon M^V \to \mathbb{R}$ by

$$J^\epsilon(u) := \frac{1}{2} \sum_{i\in V_k} d^2(u_i, u_i^n) + \lambda \sum_{(i,j)\in E} \sqrt{d^2(u_i, u_j) + \epsilon^2}. \tag{2.3}$$

Although the functional J^ϵ is differentiable the performance of standard methods to find its minimizer is poor.

[1] With minor modifications the analysis in this chapter can also be done for the TV defined by

$$TV(u) := \sum_{i\in V} \sqrt{\sum_{\substack{j\in V \\ (i,j)\in E}} d^2(u_i, u_j)}.$$

If, for example, V is a two-dimensional grid and E is the set of all pairs $(i, j) \in V \times V$ such that $j \in V$ is either to the right or to the bottom of $i \in V$ the value $\sqrt{\sum_{j\in s(i)} d^2(u_i, u_j)}$ is the 2-norm of a discrete gradient at $i \in V$. The corresponding functional is known as the isotropic total variation.

Related work

Since the beginning of DT-MRI in the 1980s, the regularization of DT-MRI images has gained interest. In DT-MRI, the values of the images are symmetric positive definite matrices. There are several proposals how to regularize such images [16, 17, 35, 51, 47]. In [30], Lellmann et al. presented a first framework and an algorithmic solution for TV regularization for arbitrary Riemannian manifolds. Their idea is to reformulate the variational problem as a multi-label optimization problem which can then be solved approximately by convex relaxation techniques. They mentioned that "Many properties of minimizers of TV-regularized models in manifolds are still not fully understood". In [53], Weinmann et al. propose a proximal point algorithm to minimize the TV functional and prove convergence for data taking values in Hadamard spaces (see Definition 2.3.3). Convergence results for spaces which are not Hadamard is an open problem we will address in this chapter.

Overview of chapter

In Section 2.1 we propose the iteratively reweighted minimization (IRM) algorithm. The purpose of the later sections is to study the convergence properties of the IRM algorithm. First, we will show a general result (Proposition 2.2.4), namely that the IRM algorithm generates a sequence for which the value of the regularized TV functional J^ϵ defined in (2.3) is non-increasing. Furthermore, under the assumption that J^ϵ has a unique critical point we prove that the sequence generated by IRM converges to that critical point (Theorem 4.2.2). For M being a Hadamard manifold, we show that these assumptions are satisfied (Proposition 2.3.1) and that the minimizer of J^ϵ converges to a minimizer of J if ϵ tends to zero (Theorem 2.3.2). Using a result from differential topology, we show that if M is a half-sphere, the functional of the optimization problem occurring in the IRM algorithm has a unique critical point (Theorem 2.4.2). Hence, IRM can also be applied to the case of M being a half-sphere. This result is of independent interest and is a first step toward a theory of convergence for spaces which are not Hadamard. Next, we will prove local linear convergence of the sequence generated by the IRM algorithm for a simple artificial image (Propositions 2.5.1 and 2.5.2). To our knowledge, for TV minimization there is no algorithm known with this property. Finally, we apply IRM and the proximal point algorithm introduced in [53] to several denoising and inpainting tasks and compare the performance in Matlab. Due to the linear convergence, we will see that our algorithm outperforms the proximal point algorithm.

Building upon our work, a C++ template library for the minimization of the TV-functional using IRM and proximal point was implemented in [19] and experiments as those in Section 2.6 were conducted.

2.1 Iteratively reweighted minimization

In this section we propose an adaptation of the well-known iteratively reweighted least squares (IRLS) algorithm. IRLS has been proved to be very successful for recovering sparse signals [18] and was already applied to scalar TV minimization problems [44]. Defining

$$w_{i,j} = (w(u))_{i,j} := (d^2(u_i, u_j) + \epsilon^2)^{-\frac{1}{2}} \text{ for all } (i,j) \in E, \tag{2.4}$$

and

$$J_w(u) := \frac{1}{2} \sum_{i \in V_k} d^2(u_i, u_i^n) + \frac{1}{2} \lambda \sum_{(i,j) \in E} w_{i,j} d^2(u_i, u_j), \tag{2.5}$$

we seek to minimize J^ϵ by alternating between reweighting and minimization steps. In a reweighting step we update the weights $w_{i,j}$ as defined in (2.4). In a minimization step we minimize the functional $J_w(u)$ defined in (2.5) with respect to u, i.e.

$$u^{new} \in \operatorname*{arg\,min}_{u \in M^V} J_w(u). \tag{2.6}$$

In the linear case, i.e. if $M = \mathbb{R}^n$ and d is the Euclidean distance, the functional J_w is a quadratic functional and can be minimized by solving a system of linear equations. This is where the term "least squares" in IRLS comes from. In the nonlinear case we propose the Riemannian Newton method (0.1). We observed that only a few steps are necessary to get a reasonable approximation of the minimizer of J_w. We call the resulting procedure the iteratively reweighted minimization (IRM) algorithm. A pseudocode description can be found below in Algorithm 1.

2.2 A general convergence result for IRM

In this section we present first results concerning the convergence of sequences generated by the IRM algorithm. These results are general in the sense that they are independent of the Riemannian manifold M. They depend, however, on Assumptions 2.2.1 and 2.2.2, which are not satisfied by every Riemannian manifold but, as we will see in Section 2.3, are satisfied by Hadamard manifolds, i.e. Riemannian manifolds with non-positive (sectional) curvature.

Assumption 2.2.1. *The functional J_w defined in* (2.5) *has a unique critical point for every $w \in \mathbb{R}^E$.*

Note that every minimizer of J_w is also a critical point of J_w and that J_w has at least one minimizer. Therefore, if Assumption 2.2.1 holds the unique critical point of J_w is also the unique minimizer of J_w. To prove convergence of a sequence generated by the IRM algorithm we will, as in [18], first reinterpret our algorithm as an alternating minimization algorithm. A first consequence will be that the values of J^ϵ are non-increasing (Proposition 2.2.4). As explained in [43], alternating minimization can fail to

Algorithm 1 Iteratively reweighted minimization

Input: Graph (V, E), noisy image $u^n \in M^{V_k}$ where $V_k \subset V$ and parameters $\lambda, \epsilon, tol > 0$.
Output: Approximation of a minimizer $u \in M^V$ of J^ϵ

```
Choose a first guess u^(0) for u (e.g. by interpolation of u^n)
Set k = 0
repeat
  w_{i,j} = (d^2(u_i^(k), u_j^(k)) + ε^2)^(-1/2) for all (i,j) ∈ E
  v^(0) = u^(k)
  Set l = 0
  repeat
    solve Hess J_w(v^(l))x = grad J_w(v^(l)) for x ∈ T_{v^(l)} M^V
    v^(l+1) = exp_{v^(l)}(-x)
    l = l + 1
  until d(v^(l), v^(l-1)) < tol
  u^(k+1) = v^(l)
  k = k + 1
until d(u^(k), u^(k-1)) < tol
return u^(k).
```

converge to a minimizer. However, using the additional Assumption 2.2.2, we will be able to prove convergence with a compactness argument (Theorem 2.2.3).

Assumption 2.2.2. *The functional J^ϵ defined in* (2.3) *has a unique critical point.*

The main goal of this section is to prove the following convergence result.

Theorem 2.2.3. *If Assumptions 2.2.1 and 2.2.2 hold, any sequence $\left(u^{(j)}\right)_{j\in\mathbb{N}}$ generated by the IRM algorithm (Algorithm 1) converges to the unique minimizer of J^ϵ.*

Theorem 2.2.3 guarantees that the sequence generated by IRM converges to the unique minimizer of J^ϵ. However, we are interested in a minimizer of J and it is at the moment not clear if the unique minimizer of J^ϵ is close to a minimizer of J for ϵ small. Furthermore, it is not clear when Assumptions 2.2.1 and 2.2.2 hold. These problems will be addressed later in Section 2.3.

2.2.1 Proof of the convergence result

As already mentioned, we will first reinterpret our algorithm as an alternating minimization algorithm. Note that the reweighting (2.4) and minimization (2.6) step of IRM are equivalent to

$$w^{new} = \left((d^2(u_i, u_j) + \epsilon^2)^{-\frac{1}{2}}\right)_{(i,j)\in E}$$

$$
\begin{aligned}
&= \arg\min_{w\in\mathbb{R}^E} \frac{1}{2}\lambda \sum_{(i,j)\in E} w_{i,j}(d^2(u_i,u_j)+\epsilon^2) + w_{i,j}^{-1} \\
&= \arg\min_{w\in\mathbb{R}^E} \frac{1}{2}\sum_{i\in V_k} d^2(u_i,u_i^n) + \frac{1}{2}\lambda \sum_{(i,j)\in E} w_{i,j}(d^2(u_i,u_j)+\epsilon^2) + w_{i,j}^{-1} \text{ and} \\
u^{new} &= \arg\min_{u\in M^V} \frac{1}{2}\sum_{i\in V_k} d^2(u_i,u_i^n) + \frac{1}{2}\lambda \sum_{(i,j)\in E} w_{i,j} d^2(u_i,u_j) \\
&= \arg\min_{u\in M^V} \frac{1}{2}\sum_{i\in V_k} d^2(u_i,u_i^n) + \frac{1}{2}\lambda \sum_{(i,j)\in E} w_{i,j}(d^2(u_i,u_j)+\epsilon^2) + w_{i,j}^{-1}.
\end{aligned}
$$

Hence, IRM is equivalent to alternating minimization of the functional

$$
\bar{J}^\epsilon(w,u) := \frac{1}{2}\sum_{i\in V_k} d^2(u_i,u_i^n) + \frac{1}{2}\lambda \sum_{(i,j)\in E} w_{i,j}(d^2(u_i,u_j)+\epsilon^2) + w_{i,j}^{-1}. \tag{2.7}
$$

A first consequence is that the value of J^ϵ is non-increasing.

Proposition 2.2.4. *Let $\left(u^{(j)}\right)_{j\in\mathbb{N}} \subset M^V$ be a sequence generated by the IRM algorithm. Then the sequence $\left(J^\epsilon(u^{(j)})\right)_{j\in\mathbb{N}} \subset \mathbb{R}$ is non-increasing.*

Proof. Note that $J^\epsilon(u) = \bar{J}^\epsilon(w(u),u)$ where J^ϵ, $\bar{J}^\epsilon$ and $w(u)$ are defined in (2.3), (2.7) and (2.4), respectively . Hence

$$
J^\epsilon(u^{(j)}) = \bar{J}^\epsilon(w(u^{(j)}),u^{(j)}) \geq \bar{J}^\epsilon(w(u^{(j)}),u^{(j+1)}) \geq \bar{J}^\epsilon(w(u^{(j+1)}),u^{(j+1)}) = J^\epsilon(u^{(j+1)}). \quad \square
$$

To prove the main result, we will need the following elementary topological lemma.

Lemma 2.2.5. *Let A be a compact space, B be a topological space, $f\colon A\to B$ be a continuous function, $a\in A$ such that $f(x)=f(a)$ if and only if $x=a$, and $(x^{(i)})_{i\in\mathbb{N}}\subset A$ be a sequence with $\lim_{i\to\infty} f(x^{(i)}) = f(a)$. Then $\lim_{i\to\infty} x^{(i)} = a$.*

Proof. Assume that there exists an open neighborhood $N(a)$ of a such that, for infinitely many $i\in\mathbb{N}$ we have $x^{(i)}\notin N(a)$. As $A\backslash N(a)$ is compact, there exists a subsequence $\left(x^{(n_i)}\right)_{i\in\mathbb{N}}$ which converges to some $\bar{a}\in A\backslash N(a)$. We now have $f(\bar{a}) = \lim_{i\to\infty} f(x^{(n_i)}) = f(a)$, which contradicts the assumption. $\square$

We are now able to prove our main theorem.

Proof of Theorem 2.2.3. Note that $\bar{J}^\epsilon(w(u^{(j)}),u^{(j)})$ is bounded from below by zero and by Proposition 2.2.4 non-increasing with j. Therefore, $\tilde{J}(w(u^{(j)}),u^{(j)})$ converges to some

value $c \in \mathbb{R}$. Note that $w(u^{(j)}) \leq \epsilon^{-1}$ for all $j \in \mathbb{N}$ and the sequence $(u^{(j)})_{j\in\mathbb{N}}$ is bounded. Hence, it has a subsequence $(u^{(n_j)})_{j\in\mathbb{N}}$ converging to some $\bar{u} \in M^V$. Let

$$u' := \operatorname*{arg\,min}_{u\in M^V} \bar{J}^\epsilon(w(\bar{u}), u).$$

By continuity of $\bar{J}^\epsilon$ we have

$$\begin{aligned} c &= \lim_{j\to\infty} \bar{J}^\epsilon(w(u^{(n_j+1)}), u^{(n_j+1)}) \\ &= \bar{J}^\epsilon(w(u'), u') \\ &\leq \bar{J}^\epsilon(w(\bar{u}), \bar{u}) \\ &= \lim_{j\to\infty} \bar{J}^\epsilon(w^\epsilon(u^{(n_j)}), u^{(n_j)}) \\ &= c. \end{aligned}$$

As we have equality and the functional $u \mapsto \bar{J}^\epsilon(w(\bar{u}), u)$ has a unique minimizer we have $u' = \bar{u}$. It follows that $(w(\bar{u}), \bar{u})$ is a critical point of $\bar{J}^\epsilon$. Therefore, $\bar{u}$ is also a critical point of J_w, with $w = w(\bar{u})$. Hence, the critical point of J^ϵ. Finally, by Proposition 2.2.4 and Lemma 2.2.5 we get that $\lim_{j\to\infty} u^{(j)} = \bar{u}$. □

2.3 IRM on Hadamard manifolds

In Theorem 2.2.3 we assume that Assumptions 2.2.1 and 2.2.2 hold, i.e. that J^ϵ and J_w for every $w \in \mathbb{R}^E$ have unique critical points. In this section we consider Hadamard manifolds, i.e. Riemannian manifolds with non-positive (sectional) curvature, for which these assumptions are satisfied.

Proposition 2.3.1. *If M is a Hadamard manifold and (V, E) is a connected graph the functional J is convex and Assumptions 2.2.1 and 2.2.2 hold.*

Furthermore we can show that the unique minimizer of J^ϵ converges to a minimizer of J if ϵ tends to zero. This shows that if we choose ϵ small the result of the IRM algorithm will be close to a minimizer of J. Note that if V_k is a proper subset of V, the functional J, unlike J^ϵ, does in general not have a unique minimizer. If, for example, $i \in V\backslash V_k$ has only two neighbors $j_1, j_2 \in V$, we can replace u_i with any value on a length-minimizing curve between u_{j_1} and u_{j_2} without changing the value of $J(u)$.

Theorem 2.3.2. *Let M be a Hadamard manifold, (V, E) connected and u^ϵ the unique minimizer of J^ϵ. Then $\lim_{\epsilon\to 0} u^\epsilon$ is well-defined and a minimizer of J.*

2.3.1 Generalization to Hadamard spaces

In fact we will prove Proposition 2.3.1 and Theorem 2.3.2 more generally for the wider class of Hadamard spaces.

Definition 2.3.3. A complete metric space (X, d) is called a Hadamard space if, for all $x_0, x_1 \in X$, there exists $y \in X$ such that, for all $z \in X$, we have

$$d^2(z, y) \leq \frac{1}{2}d^2(z, x_0) + \frac{1}{2}d^2(z, x_1) - \frac{1}{4}d^2(x_0, x_1). \tag{2.8}$$

It follows from the definition that y has to be the midpoint of x_0 and x_1 (i.e. $y \in X$ has to be such that $d(x_0, y) = d(y, x_1) = d(x_0, x_1)/2$). A comprehensive introduction to the nowadays well-established theory of Hadamard spaces is [7]. Hadamard manifolds are Hadamard spaces [14]. Note that if X is a Hadamard space then X^V with the metric

$$d(x, y) := \sqrt{\sum_{i \in V} d^2(x_i, y_i)} \quad \text{for all} \quad x,\ y \in X^V,$$

is also a Hadamard space. For any two points in a Hadamard space there exists a unique geodesic, i.e. a curve satisfying (1.1) locally, connecting them. We also need the notion of convexity on Hadamard spaces.

Definition 2.3.4. A function $f\colon X \to \mathbb{R}$ where X is a Hadamard space is called *convex* (respectively *strictly convex*) if for every nonconstant geodesic $\gamma\colon [0, 1] \to X$ (see Definition 1.1.2) the function $f \circ \gamma\colon [0, 1] \to \mathbb{R}$ is convex (respectively strictly convex).

2.3.2 Proof of Convexity of the functionals

We can now prove Proposition 2.3.1.

Proof of Proposition 2.3.1. We will first prove that the functionals J, J^ϵ and J_w are convex. The distance function (and consequently the squared distance function) is convex (see Corollary 2.5 in [49] or Proposition 5.4. in [7]). Therefore, J and J_w are convex. The convexity of J^ϵ follows by additionally using Lemma 2.3.5. We now prove strict convexity of J^ϵ and J_w. For this it is enough to show that for any $u^1, u^2 \in M^V$ with $u^1 \neq u^2$ and the geodesic $\gamma\colon [0, 1] \to M^V$ connecting u^1 and u^2, there is one term $T\colon M^V \to \mathbb{R}$ of the corresponding functional for which $T \circ \gamma$ is strictly convex. By Corollary 2.5 in [49] the function $x \mapsto d^2(x, y)$ is strictly convex for any $y \in X$. Hence, if there exists $i \in V_k$ with $u_i^1 \neq u_i^2$, we have strict convexity. If there is no $i \in V_k$ with $u_i^1 \neq u_i^2$, we can find (because (V, E) is connected) an $(i, j) \in E$ such that $u_i^1 = u_i^2$ and $u_j^1 \neq u_j^2$. Then, by the same argument as before and Lemma 2.3.5, we have strict convexity. We now prove that if M is a Hadamard manifold the functionals have a unique critical point. Since every minimizer is a critical point there exists at least one critical point. If there is more than one critical point we can connect two of them by the geodesic and get a contradiction if the values at the critical points do not coincide. Furthermore by strict convexity the values on the geodesic are strictly smaller than at the endpoints, which is a contradiction. It follows that there is only one critical point. □

To prove the convexity of J^ϵ we needed the next lemma. It states that if $f_1, \dots, f_n$ are nonnegative convex functions then $\sqrt{f_1^2 + \dots + f_n^2}$ is also convex. Note that in general the composition of two convex functions is not convex and we can therefore not use an argument like that.

Lemma 2.3.5. *If $f_1, \dots, f_n \colon [0,1] \to \mathbb{R}_{\geq 0}$ are convex functions, the function $\sqrt{f_1^2 + \dots + f_n^2}$ is also convex. Furthermore, if f_1 is strictly convex the function $\sqrt{f_1^2 + \dots + f_n^2}$ is also strictly convex.*

Proof. Note that by induction it suffices to prove the statements for $n = 2$. By convexity of f_1 and f_2 we have for any $t \in [0,1]$

$$f_1(t) \leq t f_1(1) + (1-t) f_1(0) \quad \text{and} \quad f_2(t) \leq t f_2(1) + (1-t) f_2(0).$$

Squaring (which can be done due to nonnegativity of f_1 and f_2) and adding the inequalities yields

$$(f_1^2 + f_2^2)(t) \leq t^2 (f_1^2 + f_2^2)(1) + (1-t)^2 (f_1^2 + f_2^2)(0) + 2t(1-t)(f_1(1)f_1(0) + f_2(1)f_2(0)).$$

By Cauchy–Schwarz, we get $f_1(1)f_1(0) + f_2(1)f_2(0) \leq \sqrt{f_1^2 + f_2^2}(1)\sqrt{f_1^2 + f_2^2}(0)$ and hence

$$\begin{aligned}(f_1^2 + f_2^2)(t) &\leq t^2 (f_1^2 + f_2^2)(1) + (1-t)^2 (f_1^2 + f_2^2)(0) \\ &\quad + 2t(1-t)\sqrt{f_1^2 + f_2^2}(1)\sqrt{f_1^2 + f_2^2}(0) \\ &= \left(t\sqrt{f_1^2 + f_2^2}(1) + (1-t)\sqrt{f_1^2 + f_2^2}(0) \right)^2 .\end{aligned}$$

Taking the square root shows that the function $\sqrt{f_1^2 + f_2^2}$ is convex. If f_1 is strictly convex we get strict inequality and therefore strict convexity. □

2.3.3 Proof of convergence of minimizer of J^ϵ to a minimizer of J

We can now prove the main theorem of Section 2.3. As announced, we will prove the theorem not only for manifolds M but more generally for any Hadamard space X.

Proof of Theorem 2.3.2. Let $C \subset X^V$ be the set of minimizers of J. Note that C is geodesically convex, i.e. for any geodesic $\gamma \colon [0,1] \to X^V$ with $\gamma(0), \gamma(1) \in C$ we have $\gamma(t) \in C$ for all $t \in [0,1]$. For a geodesic $\gamma \colon [0,1] \to C$ we have

$$\begin{aligned}J(\gamma(t)) &= \frac{1}{2} \sum_{i \in V_k} d^2(\gamma(t)_i, u_i^n) + \lambda \sum_{(i,j) \in E} d(\gamma(t)_i, \gamma(t)_j) \\ &\leq \frac{1}{2} \sum_{i \in V_k} t d^2(\gamma(1)_i, u_i^n) + (1-t) d^2(\gamma(0)_i, u_i^n)\end{aligned}$$

$$\begin{aligned} & +\lambda \sum_{(i,j)\in E} td(\gamma(1)_i, \gamma(1)_j) + (1-t)d(\gamma(0)_i, \gamma(0)_j) \\ = \;& tJ(\gamma(1)) + (1-t)J(\gamma(0)) \\ = \;& J(\gamma(0)) \\ \leq \;& J(\gamma(t)). \end{aligned}$$

Hence we have equality and therefore

$$d(\gamma(t)_i, \gamma(t)_j) = td(\gamma(1)_i, \gamma(1)_j) + (1-t)d(\gamma(0)_i, \gamma(0)_j) \text{ for all } (i,j) \in E \text{ and } t \in [0,1]. \tag{2.9}$$

We define $E_+ = \{(i,j) \in E | \exists u = (u_i)_{i\in V} \in C \text{ s.t. } d(u_i, u_j) > 0\}$, $E_0 = E \backslash E_+$ and the function $K\colon X^V \to \mathbb{R} \cup \{\infty\}$ by

$$K(u) := \sum_{(i,j)\in E_+} \frac{1}{d(u_i, u_j)}.$$

From (2.9) it follows that the restriction of K to C is strictly convex. Furthermore, C is compact and there exists $u \in C$ with $K(u) < \infty$. Hence there exists a unique minimizer $u^0 \in C$ of K. We define $K^\epsilon\colon X^V \to \mathbb{R}$ by

$$K^\epsilon(u) := \sum_{(i,j)\in E_+} \frac{1}{d(u_i, u_j) + \epsilon}.$$

Note that we have the inequalities

$$x + \frac{\epsilon^2}{2(x+\epsilon)} \leq \sqrt{x^2 + \epsilon^2} \leq x + \frac{\epsilon^2}{2x}.$$

Hence,

$$J(u^0) \leq J(u^\epsilon) \leq J(u^\epsilon) + \lambda\epsilon|E_0| + \frac{\lambda\epsilon^2}{2} K^\epsilon(u^\epsilon) \leq J^\epsilon(u^\epsilon) \leq J^\epsilon(u^0) \leq J(u^0) + \lambda\epsilon|E_0| + \frac{\lambda\epsilon^2}{2} K(u^0).$$

It follows that $\lim_{\epsilon\to 0} J(u^\epsilon) = J(u^0)$ and $K^\epsilon(u^\epsilon) \leq K(u^0)$. Hence,

$$d(u_i^\epsilon, u_j^\epsilon) \geq (K(u^0))^{-1} \; - \; \epsilon$$

for all $(i,j) \in E_+$. Since

$$\begin{aligned} K(u^\epsilon) \;&\geq\; K(u^0) \geq K^\epsilon(u^\epsilon) \\ &=\; K(u^\epsilon) - \epsilon \sum_{(i,j)\in E_+} \frac{1}{d(u_i^\epsilon, u_j^\epsilon)(d(u_i^\epsilon, u_j^\epsilon) + \epsilon)} \\ &\geq\; K(u^\epsilon) - \epsilon|E_+|(K(u^0))^{-2} \end{aligned}$$

we have $\lim_{\epsilon\to 0} K(u^\epsilon) = K(u^0)$. Convergence of $(u^\epsilon)_{\epsilon>0}$ to u^0 for $\epsilon \to 0$ now follows from Lemma 2.2.5 with $f\colon X^V \to \mathbb{R}^2,\ u \mapsto (J(u), K(u))$. □

2.4 IRM on the sphere

As already mentioned in the introduction the chromaticity part of an RGB-image has values on the two-dimensional sphere S^2. The m-dimensional sphere S^m together with the standard inner product is a Riemannian manifold, but not a Hadamard manifold. Therefore, we cannot apply the theory of Section 2.3. In fact even if $V = V_k$, the TV functional for $M = S^2$ does in general not have a unique minimizer as the following example shows.

Example 2.4.1. *Consider the TV functional J associated to $V = V_k = \{1, 2\}$, $E = \{(1, 2)\}$, $u_1^n = (0, 0, 1)$, $u_2^n = (0, 0, -1)$ and $\lambda > 0$. Note that any minimizer (u_1, u_2) of J satisfies $u_i \neq u_j^n$ for any $i, j \in \{1, 2\}$. By rotational symmetry of the sphere, there cannot be a unique minimizer.*

The example above is a special case where the values of the image lie opposite to each other and this raises the question if we have uniqueness if all the points lie "close" to each other. For RGB-images, we can restrict our space to

$$S^2_{\geq 0} := \left\{ x = (x_i)_{i=1}^3 \in S^2 \mid x_i \geq 0 \text{ for } i \in \{1, 2, 3\} \right\} \subset HS^2 = \left\{ x \in S^2 \,\middle|\, \sum_{i=1}^3 x_i > 0 \right\},$$

where HS^2 denotes the open half-sphere. We will show that for data on a half-sphere Assumption 2.2.1 holds. In fact, we will show a more general statement where the squared distance function $d^2(p, q)$ can be replaced with any functional of the form $\alpha(p^T q)$ where $\alpha \in C^2((-1, 1])$ is a convex and monotonically decreasing function. Note that the squared geodesic respectively the squared Euclidean distance are of this form with $\alpha(x) = \arccos^2(x)$ respectively $\alpha(x) = 2 - 2x$. The convexity of $\arccos^2$ follows from

$$(\arccos^2)'(x) = \begin{cases} \frac{-2\arccos(x)}{\sqrt{1-x^2}} & x \in (-1, 1) \\ -2 & x = 1 \end{cases} \tag{2.10}$$

$$(\arccos^2)''(x) = \begin{cases} \frac{2+(\arccos^2)'(x)x}{1-x^2} & x \in (-1, 1) \\ \frac{2}{3} & x = 1 \end{cases}, \tag{2.11}$$

and

$$(\arccos^2)''(\cos(y)) = \frac{2 - \frac{2y}{\sin(y)}\cos(y)}{\sin(y)^2} = \frac{2\cos(y)\,(\tan(y) - y)}{\sin(y)^3} \geq 0$$

for all $y \in (0, \pi)$. We can state now state the main theorem of this Section.

Theorem 2.4.2. *Let $\alpha \in C^2((-1, 1], \mathbb{R})$ be convex and monotonically decreasing, (V, E) a connected graph, $w \in \mathbb{R}^E_{>0}$, V_k a nonempty subset of V, HS^m the open half-sphere and $u^n \in (HS^m)^{V_k}$. Then the functional $J \colon (HS^m)^V \to \mathbb{R}$ defined by*

$$J(u) := \sum_{i \in V_k} \alpha(\langle u_i, u_i^n \rangle) + \sum_{(i,j) \in E} w_{(i,j)} \alpha(\langle u_i, u_j \rangle)$$

has a unique critical point.

Corollary 2.4.3. *If M is the half-sphere the Assumption 2.2.1 is satisfied, i.e. the functional J_w defined in* (2.5) *has a unique critical point for every $w \in \mathbb{R}^E$.*

We can therefore find the minimizer of J_w using the Riemannian Newton method (Section 1.3.3). Hence, we can apply IRM also for $S^2_{\geq 0}$-valued images. By Proposition 2.2.4 the functional values of J^ϵ are non-increasing. However, there are still some open questions: It is unclear if Assumption 2.2.2 holds true, i.e. if J^ϵ has a unique critical point. Furthermore, it is not clear if these minimizers converge to a minimizer of J if ϵ tends to zero. If we replace the squared distances with distances in the functional J defined in (2.2), the optimization problem becomes a multifacility location problem. This problem has been studied in [6, 21]. However, existence and uniqueness of a minimizer is still an open problem.

Without loss of generality we may assume that $w_{i,j} = w_{j,i} > 0$ for all $(i,j) \in E$ in Theorem 2.4.2. Even though J itself is not convex, we can prove that J is locally strictly convex at every critical point of J.

Lemma 2.4.4. *If u is a critical point of J we have that* Hess $J(u)$ *is positive definite.*

We will prove Lemma 2.4.4 in Section 2.4.1. To prove Theorem 2.4.2 we will also need the Poincaré–Hopf theorem, a result from differential topology.

Theorem 2.4.5 (Poincaré–Hopf [34])**.** *Let M be a compact manifold with boundary ∂M and $U\colon M \to TM$ a vector field on M such that U is pointing outward on ∂M. Assume that U has a continuous derivative DU, all zeros of U are isolated and $DU(z)$ is invertible for all zeros $z \in M$ of U. Then U has finitely many zeros $z_1, \dots, z_n \in M$ and*

$$\sum_{i=1}^{n} sgn(\det(DU(z_i))) = \chi(M),$$

where sgn denotes the sign function and $\chi(M)$ the Euler characteristic of M.

Unfortunately, the space of our interest $(HS^m)^V$, i.e. the Cartesian power of the open half-sphere, is from a differential topology viewpoint not a manifold with boundary but a manifold with corners. However, $(HS^m)^V$ can be approximated arbitrarily close by a manifold with boundary which allows us to still use the Poincaré–Hopf theorem.

To apply the Poincaré–Hopf theorem, we need to prove that the gradient of J at the boundary

$$\partial\,(HS^m)^V = \{u \in (S^m)^V \,|(u_i)_{m+1} \geq 0 \text{ for all } i \in V,\ \exists i \in V \text{ s.t. } (u_i)_{m+1} = 0\},$$

of $(HS^m)^V$ is pointing outward.

Lemma 2.4.6. *Let $u = (u_i)_{i\in V} \in \partial\,(HS^m)^V$. Then we have $(\operatorname{grad}_{u_i} J(u))_{m+1} \leq 0$ for all $i \in V$ with $u_i \in \partial HS^m$ and there exists at least one such i with $(\operatorname{grad}_{u_i} J(u))_{m+1} < 0$.*

Proof. For $i \in V$ with $u_i \in \partial HS^m$ we have

$$\begin{aligned}(\operatorname{grad}_{u_i} J(u))_{m+1} &= e_{m+1}^T \operatorname{grad}_{u_i} J(u) \\ &= e_{m+1}^T P_{T_{u_i}M} \operatorname{grad}_{u_i} \bar{J}(u) \\ &= e_{m+1}^T \operatorname{grad}_{u_i} \bar{J}(u) \\ &= 1_{V_k}(i)\alpha'(u_i^T u_i^n) e_{m+1}^T u_i^n + \sum_{j \in n(i)} w_{i,j}\alpha'(u_i^T u_j) e_{m+1}^T u_j \\ &\leq 0.\end{aligned}$$

Let $j \in V$ be such that $u_j \in \partial HS^m$. Consider a path $j_0, j_1, \ldots, j_l \in V$ from $j_0 = j$ to $j_l \in V_k$. Let i be the largest index of this path such that $u_i \in \partial HS^m$. Then we have strict inequality above. □

We are now able to prove the main result of Section 2.4.

Proof of Theorem 2.4.2. Consider the gradient vector field $\operatorname{grad} J$ of J. By Lemma 2.4.4 all zeros z are isolated and satisfy $\det(D \operatorname{grad} J(z)) > 0$. By Lemma 2.4.6, the vector field $\operatorname{grad} J$ is pointing outward at the boundary $\partial \left(HS^m\right)^V$. Hence, by Theorem 2.4.5 the number of zeros of $\operatorname{grad} J$ is $\chi\left(\left(HS^m\right)^V\right) = \chi\left(HS^m\right)^{|V|} = 1$. □

2.4.1 Convexity at critical points

The purpose of this section is to prove Lemma 2.4.4. By the computations in Section 1.5.1 the (intrinsic) Taylor expansion of $\alpha(x^T y)$ is

$$\begin{aligned}\alpha((\exp_x(r))^T \exp_y(s)) &= \alpha(x^T y) + \alpha'(x^T y)(y^T r + x^T s) \\ &\quad + \frac{1}{2}\alpha''(x^T y)(y^T r + x^T s)^2 \\ &\quad + \frac{1}{2}\begin{pmatrix} r^T & s^T \end{pmatrix} \begin{pmatrix} -\beta(x^T y) I_{m+1} & \alpha'(x^T y) I_{m+1} \\ \alpha'(x^T y) I_{m+1} & -\beta(x^T y) I_{m+1} \end{pmatrix} \begin{pmatrix} r \\ s \end{pmatrix} \\ &\quad + \mathcal{O}(|r|^3 + |s|^3),\end{aligned}$$

where $\beta \colon C^1((-1, 1])$ is defined by $\beta(x) = x\alpha'(x)$.

For $u \in \left(HS^m\right)^V$ we define the matrix $T(u) \in \mathbb{R}^{V \times V}$ by

$$T(u)_{ij} := \begin{cases} -1_{V_k}(i)\beta((u_i^n)^T u_i) - \sum_{k \in n(i)} w_{i,k}\beta(u_i^T u_k), & i = j \\ w_{i,j}\alpha'(u_i^T u_j), & (i,j) \in E \\ 0, & \text{otherwise,} \end{cases}$$

where $n(i) := \{j \in V | (i,j) \in E\}$ denotes the set of neighbors of $i \in V$ and $1_{V_k} \in \{0,1\}^V$ is the indicator function of V_k.

Lemma 2.4.7. *If the matrix $T(u)$ is positive definite the Hessian* Hess $J(u)$ *is also positive definite.*

Proof. Since $\alpha''(t) \geq 0$ for all $t \in (-1, 1]$ the term (2.12) is nonnegative. Adding up the second part (2.12) of the Hessian for all the terms in the functional J yields the matrix $T(u) \otimes I_{m+1}$ where $\otimes$ denotes the Kronecker product. □

We now prove an identity for the critical points.

Lemma 2.4.8. *Let $u \in (HS^m)^V$ be a critical point of J. Then we have*

$$-1_{V_k}(i)\alpha'(u_i^T u_i^n)u_i^n = (T(u)u)_i := \sum_{j \in V} T(u)_{ij} u_j \quad \textit{for all } i \in V.$$

Proof. Let $\bar{J}$ be the natural extension of J which is defined by

$$\bar{J}\colon \left\{ u \in \left(\mathbb{R}^{m+1}\right)^V \;\middle|\; u_i^T u_i^n \geq -1 \text{ for all } i \in V \text{ and } u_i^T u_j \geq -1 \text{ for all } (i,j) \in E \right\} \to \mathbb{R}.$$

Then, for all $i \in V$ there exists $\mu_i \in \mathbb{R}$ such that

$$\mu_i u_i = \frac{d\bar{J}}{du_i} = 1_{V_k}(i)\alpha'(u_i^T u_i^n)u_i^n + \sum_{j \in n(i)} w_{i,j}\alpha'(u_i^T u_j)u_j. \tag{2.12}$$

Multiplying equation (2.12) with u_i yields

$$\mu_i = 1_{V_k}(i)\beta(u_i^T u_i^n) + \sum_{j \in n(i)} w_{i,j}\beta(u_i^T u_j).$$

Therefore, for all $i \in V$ we have

$$\left(1_{V_k}(i)\beta(u_i^T u_i^n) + \sum_{j \in n(i)} w_{i,j}\beta(u_i^T u_j)\right) u_i = 1_{V_k}(i)\alpha'(u_i^T u_i^n)u_i^n + \sum_{j \in n(i)} w_{i,j}\alpha'(u_i^T u_j)u_j,$$

which can be rewritten in the desired form. □

We can now prove the desired result.

Proof of Lemma 2.4.4. Since HS^n is an open half-sphere there exists a vector $e \in S^n$ with $e^T w > 0$ for all $w \in HS^n$. Let $r := (e^T u_i)_{i \in V} \in \mathbb{R}^V_{>0}$. By Lemma 2.4.8 and $\alpha'(t) \leq 0$ for all $t \in (-1, 1]$ we have

$$(T(u)r)_i = e_{m+1}^T (T(u)u)_i = -1_{V_k}(i)\alpha'(u_i^T u_i^n)(u_i^n)_{m+1} \geq 0 \quad \text{for all} \quad i \in V. \tag{2.13}$$

It follows that $S := diag(r)T(u)diag(r)$, where $diag(r)$ is the diagonal matrix with entries of r on the diagonal, is diagonally dominant and therefore positive semidefinite.

We now prove that S is even positive definite. Assume that $v \in \mathbb{R}^V$ is an eigenvector of S with eigenvalue 0 and let $i \in V$ such that $|v_i| \geq |v_j|$ for all $j \in V$. As $(Sv)_i = 0$, it follows that $v_j = v_i$ for all $j \in n(i)$ and recursively that $v_j = v_i$ for all $j \in V$. Let $j \in V_k$ then

$$(Sv)_j = r_j(T(u)r)_j v_j = -r_j \alpha'(u_j^T u_j^n)(u_j^n)_{m+1} v_j \neq 0,$$

which is a contradiction. Hence, S is positive definite and therefore $T(u)$ is positive definite as well. By Lemma 2.4.7, we have that $\operatorname{Hess} J(u)$ is then also positive definite. □

2.5 Linear convergence of IRM on a test image

To get some ideas of the local convergence speed, we will analyze the TV functional on a very simple artificial test image consisting of only two pixels with values 0 and 1. For this minimization problem, we can write down the exact solution and study the convergence speed of IRM and the proximal point algorithm [53]. Even though this image is not realistic, the convergence behavior of IRM and proximal point persists also for larger images with values in any Riemannian manifold.

2.5.1 The TV functional of the test image

The TV functional for the image with only two pixels with values 0 and 1 is

$$J(u_0, u_1) = \frac{1}{2}(u_0 - 0)^2 + \frac{1}{2}(u_1 - 1)^2 + \lambda |u_0 - u_1|. \tag{2.14}$$

By symmetry, the minimizer $u^0 = (u_0^0, u_1^0)$ satisfies $u_0^0 + u_1^0 = 1$. For $u = (u_0, u_1) \in \mathbb{R}^2$ let $y(u) := u_1 - u_0$. If $u_0 + u_1 = 1$, we have $J(u) = \frac{1}{4}(y(u) - 1)^2 + \lambda |y(u)|$. The minimizer u^0 satisfies

$$y(u^0) = \begin{cases} 1 - 2\lambda & \text{if } \lambda < \frac{1}{2} \\ 0 & \text{if } \lambda \geq \frac{1}{2} \end{cases},$$

i.e. the minimizer is

$$(u_0^0, u_1^0) = \begin{cases} (\lambda, 1 - \lambda) & \text{if } \lambda < \frac{1}{2} \\ (\frac{1}{2}, \frac{1}{2}) & \text{if } \lambda \geq \frac{1}{2} \end{cases}.$$

The regularized functional is

$$J^\epsilon(u_0, u_1) = \frac{1}{2}(u_0 - u_0^0)^2 + \frac{1}{2}(u_1 - u_1^0)^2 + \lambda \sqrt{(u_0 - u_1)^2 + \epsilon^2}.$$

For $u = (u_0, u_1)$ with $u_0 + u_1 = 1$, we have $J^\epsilon(u) = \frac{1}{4}(y(u) - 1)^2 + \lambda \sqrt{(y(u))^2 + \epsilon^2}$. Taking the derivative by y we get the condition $f(y(u^\epsilon)) = 0$ for the minimizer u^ϵ of J^ϵ where

$$f(y) := \frac{1}{2}(y - 1) + \frac{\lambda y}{\sqrt{y^2 + \epsilon^2}}.$$

By convexity of J^ϵ, the function f is monotone increasing. If $\lambda \leq \frac{1}{2}$, we have $f(1-2\lambda) < 0$ and hence

$$y(u^\epsilon) > 1 - 2\lambda. \tag{2.15}$$

If $\lambda > \frac{1}{2}$ we have

$$f\left(\frac{\epsilon}{\sqrt{4\lambda^2 - 1}}\right) = \frac{\epsilon}{2\sqrt{4\lambda^2 - 1}} > 0$$

and hence

$$y(u^\epsilon) < \frac{\epsilon}{\sqrt{4\lambda^2 - 1}}. \tag{2.16}$$

With some additional work, one can show that

$$|y(u^\epsilon) - y(u^0)| \leq \begin{cases} \mathcal{O}(\epsilon^2) & \lambda < \frac{1}{2} \\ \mathcal{O}(\epsilon^{\frac{2}{3}}) & \lambda = \frac{1}{2} \\ \mathcal{O}(\epsilon) & \lambda > \frac{1}{2} \end{cases}. \tag{2.17}$$

2.5.2 Proof of linear convergence with constant ϵ

We now prove that if $\lambda \neq 0.5$, the sequence generated by IRM converges linearly to u^ϵ with a convergence rate independent of $\epsilon > 0$.

Proposition 2.5.1. *Let $\lambda \in \mathbb{R}_{>0} \setminus \left\{\frac{1}{2}\right\}$, $u^{(0)} \in \mathbb{R}^2$ and $\epsilon > 0$. Then the sequence $(u^{(k)})_{k\in\mathbb{N}}$ generated by the IRM algorithm converges linearly to u^ϵ with (asymptotic) rate of convergence at most* $\min(2\lambda, (2\lambda)^{-1})$.

Proof. Note that we have $u_0^{(k)} + u_1^{(k)} = 1$ for all $k \geq 1$. Hence, it is enough to prove that $y^{(k)} := y(u^{(k)})$ converges linearly to $y(u^\epsilon)$. We have

$$w(u) = \frac{1}{\sqrt{(u_1 - u_0)^2 + \epsilon^2}} = \frac{1}{\sqrt{y(u)^2 + \epsilon^2}}$$

and

$$y^{(k+1)} = \frac{1}{1 + 2\lambda w(u^{(k)})} = \frac{1}{1 + 2\lambda\left(\left(y^{(k)}\right)^2 + \epsilon^2\right)^{-\frac{1}{2}}} = G_\epsilon(y^{(k)}),$$

where

$$G_\epsilon(y) := \frac{1}{1 + 2\lambda(y^2 + \epsilon^2)^{-\frac{1}{2}}} = \frac{(y^2 + \epsilon^2)^{\frac{1}{2}}}{(y^2 + \epsilon^2)^{\frac{1}{2}} + 2\lambda}. \tag{2.18}$$

By Theorem 2.2.3 the sequence $(u^{(k)})_{k\in\mathbb{N}}$ and therefore $\left(y^{(k)}\right)_{k\in\mathbb{N}}$ converges to u^ϵ respectively $y(u^\epsilon)$. Let $y^\epsilon = y(u^\epsilon)$. The (asymptotic) convergence rate is given by the absolute value of the derivative of G_ϵ at y^ϵ. We have

$$|G'_\epsilon(y)| = \left|\frac{2\lambda y}{(y^2 + \epsilon^2)^{\frac{1}{2}}((y^2 + \epsilon^2)^{\frac{1}{2}} + 2\lambda)^2}\right|$$

$$\begin{aligned} &= \left|\frac{y}{(y^2+\epsilon^2)^{\frac{1}{2}}}\right|\left|\frac{2\lambda}{((y^2+\epsilon^2)^{\frac{1}{2}}+2\lambda)^2}\right| \\ &< \left|\frac{2\lambda}{((y^2+\epsilon^2)^{\frac{1}{2}}+2\lambda)^2}\right|. \end{aligned}$$

Therefore we have

$$\left|G'_\epsilon(y)\right| < \frac{2\lambda}{(2\lambda)^2} = (2\lambda)^{-1} \text{ for all } y \geq 0. \tag{2.19}$$

For $\lambda < \frac{1}{2}$ we have by Equation (2.15)

$$\left|G'_\epsilon(y^\epsilon)\right| < \frac{2\lambda}{(y^\epsilon+2\lambda)^2} < 2\lambda. \quad \square$$

2.5.3 Proof of linear convergence in the case of ϵ converging to zero

An interesting adaptation of our algorithm is to decrease $\epsilon > 0$ during the algorithm. Even though the analysis so far is only performed for a fixed $\epsilon > 0$, we believe that the sequence of the adapted algorithm converges linearly to a minimizer of J. For our simple test image we can prove this statement.

Proposition 2.5.2. *Let $\lambda \in \mathbb{R}_{>0} \setminus \left\{\frac{1}{2}\right\}$, $u^{(0)} \in \mathbb{R}^2$, u^0 be the minimizer of J defined in (2.14) and $(\epsilon^{(k)})_{k\in\mathbb{N}} \subset \mathbb{R}_{>0}$ be a sequence that converges linearly to 0 with convergence rate smaller than $\min(\sqrt{2\lambda}, (2\lambda)^{-1})$. Then the sequence $(u^{(k)})_{k\in\mathbb{N}}$ generated by the IRM algorithm where for the k-th reweighting step ϵ is replaced with $\epsilon^{(k)}$ satisfies*

$$\left|u^{(k)} - u^0\right| \leq C_1 C_2^k.$$

for some $C_1 > 0$ and $C_2 \in (0,1)$.

Proof. Let $y^{(k)} = y(u^{(k)})$. It is enough to prove that there exists $C_1 > 0$ and $C_2 \in (0,1)$ with

$$\left|y^{(k)} - y^0\right| \leq C_1 C_2^k. \tag{2.20}$$

Note that

$$y^{(k+1)} = G_{\epsilon^{(k)}}(y^{(k)}),$$

where G_ϵ is defined in (2.18). Assume first $\lambda > \frac{1}{2}$ and therefore $y^0 = 0$. By (2.19) and (2.16), we have

$$\begin{aligned} y^{(k+1)} &\leq \left|y^{(k+1)} - y^{\epsilon^{(k)}}\right| + y^{\epsilon^{(k)}} \\ &\leq \frac{1}{2\lambda}\left|y^k - y^{\epsilon^{(k)}}\right| + \frac{\epsilon^{(k)}}{\sqrt{4\lambda^2-1}} \\ &\leq \frac{1}{2\lambda} y^k + \epsilon^{(k)}\left(\frac{1}{2\lambda}+1\right)\frac{1}{\sqrt{4\lambda^2-1}}. \end{aligned}$$

Using this inequality iteratively yields

$$y^{(k+1)} \leq \frac{1}{(2\lambda)^k} y^{(1)} + \left(\epsilon^{(k)} + \frac{1}{2\lambda} \epsilon^{(k-1)} + \cdots + \frac{1}{(2\lambda)^{k-1}} \epsilon^{(1)} \right) \left(\frac{1}{2\lambda} + 1 \right) \frac{1}{\sqrt{4\lambda^2 - 1}}.$$

Since $(\epsilon^{(k)})_{k \in \mathbb{N}} \subset \mathbb{R}_{>0}$ converges linearly to 0 with convergence rate smaller than $(2\lambda)^{-1}$ there exists $C_3 > 0$ and $C_4 < (2\lambda)^{-1}$ such that we have $\epsilon^{(k)} \leq C_3 C_4^k$. Hence, we have

$$\begin{aligned} y^{(k+1)} &\leq \frac{1}{(2\lambda)^k} y^{(1)} + \left(\epsilon^{(k)} + \frac{1}{2\lambda} \epsilon^{(k-1)} + \cdots + \frac{1}{(2\lambda)^{k-1}} \epsilon^{(1)} \right) \left(\frac{1}{2\lambda} + 1 \right) \frac{1}{\sqrt{4\lambda^2 - 1}} \\ &\leq \left(\frac{1}{2\lambda} \right)^k \left(y^{(1)} + C_3 \left(\frac{1}{2\lambda} + 1 \right) \frac{1}{\sqrt{4\lambda^2 - 1}} \frac{C_4 2\lambda}{1 - C_4 2\lambda} \right) \end{aligned}$$

which shows that (2.20) holds. Let now $\lambda < \frac{1}{2}$ and therefore $y^0 = 1 - 2\lambda$. Note that $y^{(1)} > 0$. If $\delta > 0$ and $y^{(k)} < 1 - 2\lambda - \delta$, we have

$$y^{(k+1)} = G_{\epsilon^{(k)}}(y^{(k)}) > \frac{y^{(k)}}{y^{(k)} + 2\lambda} > \frac{y^{(k)}}{1 - \delta}.$$

Furthermore if $y^{(k)} > 1 - 2\lambda$, it follows

$$y^{(k+1)} = G_{\epsilon^{(k)}}(y^{(k)}) > \frac{y^{(k)}}{y^{(k)} + 2\lambda} > \frac{1 - 2\lambda}{(1 - 2\lambda) + 2\lambda} = 1 - 2\lambda.$$

Hence we have $y^{(k)} > 1 - 2\lambda - \delta$ for all $k \in \mathbb{N}$ large enough. For $C_5 \in (2\lambda, 1)$ and $0 < \delta < \min\left(1 - 2\lambda, 1 - \sqrt{\frac{2\lambda}{C_5}}\right)$ we have for $y > 1 - 2\lambda - \delta$, that

$$|G'_\epsilon(y)| < \frac{2\lambda}{(y + 2\lambda)^2} < \frac{2\lambda}{(1 - \delta)^2} < C_5.$$

Therefore, we have

$$\left| y^{(k+1)} - y^{\epsilon^{(k)}} \right| = \left| G_{\epsilon^{(k)}}(y^{(k)}) - G_{\epsilon^{(k)}}(y^{\epsilon^{(k)}}) \right| \leq C_5 \left| y^{(k)} - y^{\epsilon^{(k)}} \right|.$$

By (2.17), there exists $C_6 > 0$ such that $\left| y^{\epsilon^{(k)}} - y^0 \right| < C_6 \left(\epsilon^{(k)} \right)^2$. Hence, we have

$$\begin{aligned} \left| y^{(k+1)} - y^0 \right| &\leq \left| y^{(k+1)} - y^{\epsilon^{(k)}} \right| + \left| y^{\epsilon^{(k)}} - y^0 \right| \\ &\leq C_5 \left| y^{(k)} - y^{\epsilon^{(k)}} \right| + C_6 \left(\epsilon^{(k)} \right)^2 \\ &\leq C_5 \left| y^{(k)} - y^0 \right| + \left(\epsilon^{(k)} \right)^2 C_6 (1 + C_5). \end{aligned}$$

The proof can now be finished with the same argument as in the case $\lambda > \frac{1}{2}$. □

2.5.4 Comparison with convergence speed of proximal point algorithm

We will now compare the convergence speed of our method with the convergence speed of the proximal point algorithm proposed in [53] (see Algorithm 2).

The proximal point method

To motivate the proximal point method consider a convex differentiable functional $J\colon \mathbb{R}^n \to \mathbb{R}$. To compute the minimizer of J we can solve the ordinary differential equation (ODE)

$$\dot{u} = -\operatorname{grad} J(u).$$

The minimizer of J is then given by $\lim_{t\to\infty} u(t)$. To numerically solve the ODE we can perform implicit Euler steps with step sizes $(\mu_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{>0}$, i.e. we define $(u_k)_{k\in\mathbb{N}} \subset \mathbb{R}^n$ recursively as the solution of

$$u_{k+1} = u_k - \mu_k \operatorname{grad} J(u_{k+1}). \tag{2.21}$$

This method can be generalized to (non-differentiable) functionals on a metric space X by replacing the implicit Euler step (2.21) with $u_{k+1} := prox_{\mu_k J}(u_k)$ where the proximal map $prox$ is defined below.

Definition 2.5.3. For (X, d) a metric space and $f\colon X \to \mathbb{R}$, the proximal map $prox_f\colon X \to X$ is defined by

$$prox_f(x) := \operatorname*{arg\,min}_{y\in X} f(y) + \frac{1}{2}d^2(x, y).$$

The proximal point algorithm makes use of the fact that if V is a two-dimensional grid and X a geodesic space the TV functional J defined in (2.2) can be written as a sum of functionals for which there exists an explicit formula for the proximal map in terms of geodesics. We now prove the negative result that for the proximal point algorithm we do in general not have linear convergence.

Sublinear convergence for proximal point algorithm

We can now prove that a sequence generated by the proximal point algorithm, unlike a sequence generated by the IRM algorithm, does in general **not** convergence linearly to u^0.

Proposition 2.5.4. *Let $\lambda < \left(0, \frac{1}{2}\right)$. A sequence $(u^{(k)})_{k\in\mathbb{N}}$ generated by the proximal point algorithm does* **not** *converge linearly to u^0.*

Proof. Let $(\mu_k)_{k\in\mathbb{N}} \in \ell^2\backslash\ell^1$ be the parameter sequence and $(u^{(k)})$ be the corresponding sequence generated by the proximal point algorithm. Note that we have $u_0^{(k)} + u_1^{(k)} = 1$ for all $k \in \mathbb{N}$. Let $v^{(k)} := u_0^{(k)}$. For k large enough, we have $v^{(k)} + \lambda\mu_k \leq \frac{1}{2}$ and therefore

$$v^{(k+1)} = v^{(k)} + \frac{\mu_k\lambda - \frac{\mu_k}{1+\mu_k}v_k}{5}.$$

Algorithm 2 Proximal point algorithm

Input: Noisy image $a \in X^{m\times n}$, $\lambda > 0$ and sequence $(\mu_k)_{k\in\mathbb{N}} \in \ell^2\backslash\ell^1$.

Output: Approximation for the minimizer $u \in X^{m\times n}$ of $J(a, \lambda)$

$u = a$
for $k = 1, \dots$ **do**
  for $i = 1, \dots, m$; $j = 1, \dots, n$ **do**
    $t_{ij} = \mu_k d(a_{i,j}, x_{i,j})/(1 + \mu_k)$
    $z_{i,j}^{(1)} = [u_{i,j}, a_{i,j}]_{t_{ij}}$
    $t_{ij} = \min(\lambda\mu_k, d(u_{i,j}, u_{i,j+1})/2)$
    $z_{i,j}^{(2)} = [u_{i,j}, u_{i,j+1}]_{t_{ij}}$
    $t_{ij} = \min(\lambda\mu_k, d(u_{i,j}, u_{i,j-1})/2)$
    $z_{i,j}^{(3)} = [u_{i,j}, u_{i,j-1}]_{t_{ij}}$
    $t_{ij} = \min(\lambda\mu_k, d(u_{i,j}, u_{i+1,j})/2)$
    $z_{i,j}^{(4)} = [u_{i,j}, u_{i+1,j}]_{t_{ij}}$
    $t_{ij} = \min(\lambda\mu_k, d(u_{i,j}, u_{i-1,j})/2)$
    $z_{i,j}^{(5)} = [u_{i,j}, u_{i-1,j}]_{t_{ij}}$
    $u'_{i,j} = approx\_mean(z_{ij}^{(1)}, z_{ij}^{(2)}, z_{ij}^{(3)}, z_{ij}^{(4)}, z_{ij}^{(5)})$
  end for
  for $i = 1, \dots, m$; $j = 1, \dots, n$ **do**
    $u_{i,j} = u'_{i,j}$
  end for
end for

Here $c = [a, b]_t \in X$ is the unique value on the geodesic between a and b with $d(a, c) = t$ and *approx_mean* is an approximative variant of the Riemannian average.

Hence

$$v^{(k+1)} - \lambda = \left(1 - \frac{\mu_k}{5(1+\mu_k)}\right)\left(v^{(k)} - \lambda\right) + \frac{\mu_k^2 \lambda}{5(1+\mu_k)}.$$

Assume that there exists $N_0 \in \mathbb{N}$ and $C_1 > 0$, $0 < C_2 < 1$ with $\left|v^{(k)} - \lambda\right| \leq C_1 C_2^k$ for all $k > N_0$. Then we have

$$\mu_k \leq \sqrt{\frac{5(1+\mu_k)}{\lambda}\left(1 - \frac{\mu_k}{5(1+\mu_k)} + C_2\right) C_1 C_2^k} \leq C_3 \left(\sqrt{C_2}\right)^k,$$

and therefore $(\mu_k)_{k\in\mathbb{N}} \in \ell^1$, which is a contradiction. □

Comparison of convergence speed in a numerical test

We compared the two algorithms for the simple test image also with a numerical experiment. In Figure 2.1, we plotted the error $e(u) := |u_0 - u_0^0|$ for $\lambda = 0.15$ and $\epsilon^{(k)} = 10^{-k}$ in dependence of the number of iterations for both algorithms. As we can see for the IRM algorithm we have linear convergence, whereas for the proximal point algorithm the convergence is much slower and asymptotically not linear. Hence, in order to compute the TV minimizer of an image up to high precision IRM is more appropriate than proximal point. However, in most applications we do not have to compute u^0 up to high precision and it is sufficient to use about $3-5$ IRM or $10-100$ proximal point steps.

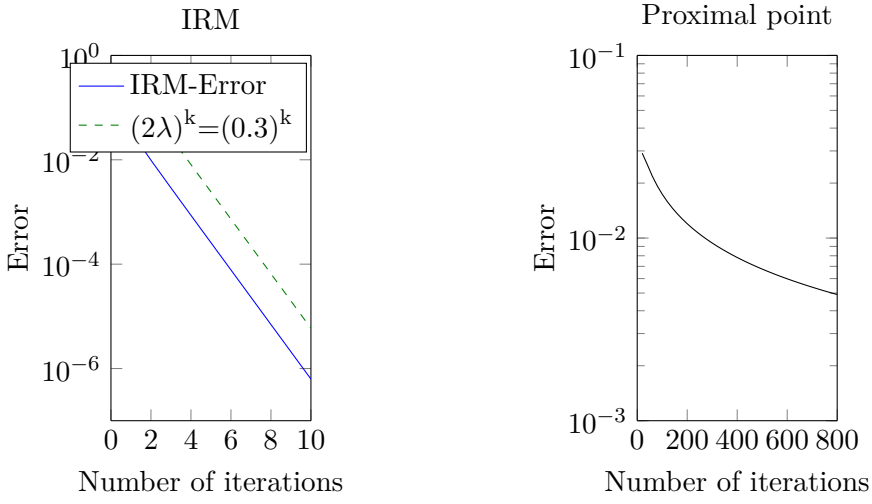


Figure 2.1: Error in dependence of the number of iterations for $\lambda = 0.15$. For the IRM algorithm we choose $\epsilon^{(k)} = 10^{-k}$ and for the proximal point algorithm the sequence $\mu_k = 3k^{-0.95}$.

2.6 Numerical experiments

We start with a few remarks regarding the implementation of IRM (Algorithm 1). The initial guess $u^{(0)}$ cannot be chosen arbitrarily. The reason is that the Newton algorithm converges only locally. However, as observed in practice, a simple smoothing filter yields a first guess for which the algorithm converges.

If our graph (V, E) is sparse, the Hessian of the functional J_w defined in (2.5) will be sparse as well, which allows us to solve the linear system of a Newton iteration in moderate time. We used a direct solver. Conjugate gradients can also be used: it is, however, observed that the unconditioned version is slightly slower. With a suitable preconditioner, it could be possible to improve the convergence speed.

For Euclidean spaces the functional J_w^ϵ is quadratic and the minimization problem (2.6) boils down to a linear system of equations. Hence for Euclidean data, we can restrict the number of Newton iterations in each IRM step to one. In practice, if we have manifold-valued data and use only one Newton iteration in each IRM step, the IRM algorithm still converges. However, we do not have a theory which proves convergence in this case. An option to choose the stopping criteria, for which we can guarantee convergence while reducing the computational cost, is to do Newton iterations until the value of J^ϵ is smaller than before the first Newton iteration (which usually happens after one Newton iteration).

To apply Algorithm 1 we used the computations of the gradient and the Hession of the squared distance function on S^m, $SPD(n)$ and $SO(n)$ from Section 1.5. The numerical experiments were conducted on a laptop using a single 2.9 GHz Intel Core i7 processor and the Matlab numerical computing environment.

For larger images it is not recommended to use IRM directly since this would require a large amount of memory and computational time. Instead, one could divide the image into smaller subimages, apply the algorithm to each of these subimages and finally compose the denoised subimages again to a complete denoised image.

2.6.1 Sphere-valued images

In Section 2.6.1 we explain how the second derivative of the squared spherical distance can be computed using a result from Section 1.5.1. In the following we present two applications with sphere-valued data.

Computations

From Section 1.5.1, we have

$$d^2(\exp_x(r), \exp_y(s)) \quad = \quad d^2(x, y) + \alpha'(y^T r + x^T s)$$

$$+\frac{1}{2}\begin{pmatrix} r^T & s^T \end{pmatrix}\left(\alpha''\begin{pmatrix} yy^T & yx^T \\ xy^T & xx^T \end{pmatrix}+\begin{pmatrix} -\beta I & \alpha' I \\ \alpha' I & -\beta I \end{pmatrix}\right)\begin{pmatrix} r \\ s \end{pmatrix}$$
$$+\mathcal{O}(|r|^3+|s|^3),$$

for any $x, y \in S^m$, $r \in T_xM$ and $s \in T_yM$, where $\alpha = \alpha(x^Ty) = \arccos^2(x^Ty)$ and $\beta = \beta(x^Ty) := x^Ty\alpha'(x^Ty)$. The derivatives of $\arccos^2$ are given in (2.10). The sum of the two matrices given in (2.22) defines not only a linear map $T_xS^m \times T_yS^m \to T_xS^m \times T_yS^m$, but a linear map $\mathbb{R}^{m+1} \times \mathbb{R}^{m+1} \to \mathbb{R}^{m+1} \times \mathbb{R}^{m+1}$. To restrict ourselves to the tangent space at $(x, y) \in S^m \times S^m$, orthogonal bases of T_xS^m and T_yS^m using QR-decomposition are constructed. By a change of basis, we compute the gradient and Hessian of d^2 with respect to the new basis. Therefore, we can compute second derivatives of squared distance functions and hence solve the optimization problems (2.6). The IRM algorithm for TV regularization on spheres, following the explanations above, was implemented in [15].

Inpainting

In Figure 2.2, we see an example of color inpainting. We first detect the region to inpaint (the blue lines). Next, we do a scattered interpolation to get a first guess of our image. Finally, we apply our TV minimization algorithm with $\lambda = 5 \cdot 10^{-3}$. As we can see, the clear straight edges occur only after we do the TV minimization.

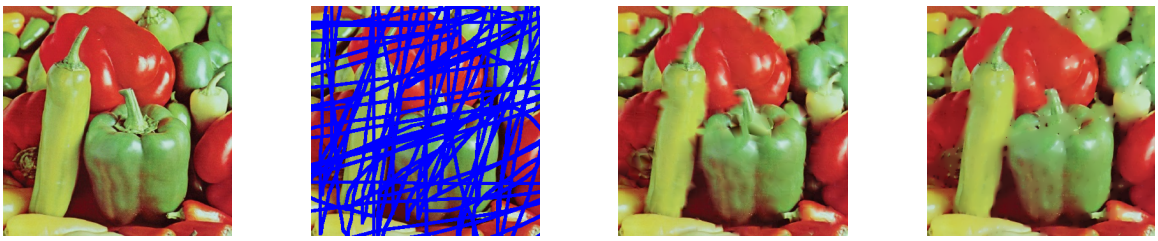


Figure 2.2: Color inpainting. From left to right: original, damaged, first guess and restored image.

Colorization

In Figure 2.3, we see an example of colorization applied to an image known among the image processing community as Lena. We assumed that the brightness is known, but the color part of every pixel is only known with probability 0.01. We first detected the edges from the grayscale image using the Canny edge detector [12]. Next, we computed a first approximation of the color part by a scattered interpolation. Finally, we computed the color part by minimizing a weighted (weight 0.01 at the edges and 1 everywhere else) TV functional with $\lambda = 10^{-2}$.

Figure 2.3: Colorization. From left to right: original image, image when almost all color was removed and restored image.

2.6.2 Matrix-valued images

We will consider *SPD* -valued and $SO(n)$-valued images.

SPD -valued images

By [11] (Page 314) the space of positive definite matrices is a Hadamard manifold (i.e. it has non-positive sectional curvature) and the theory of Section 2.3 can therefore be applied. The IRM algorithm for TV regularization of *SPD* -matrices, following the explanations above, was implemented in [38]. In Figure 2.4, we can see denoising of a real 32×32 DT-MRI image [8]. We choose $\lambda = 0.7$.

SO(n)-valued images

Unfortunately, $SO(n)$ is not a Hadamard manifold and we cannot apply the theory of Section 2.3. However, in practice it was nevertheless possible to do denoising and inpainting of $SO(n)$-valued images. In Figure 2.5, we can see denoising of an artificial 10×10 $SO(3)$-valued image, where for each $(i, j) \in \{1, \dots, 10\}^2$ the value of the pixel (i, j) is the rotation with axis $(i,\ j,\ 0)$ and angle $1 + (i + j)/10$. The $SO(3)$-matrices are illustrated by rotated cubes. To add noise, we added a random matrix with Gaussian distributed (standard deviation 0.3) entries and projected back to $SO(3)$. For $A \in \mathbb{R}^{n \times n}$ with singular value decomposition $A = U\Sigma V^T$, the projection onto $SO(n)$ is given by $P_{SO(n)}(A) = UV^T$, i.e. by dropping Σ. We choose $\lambda = 0.3$.

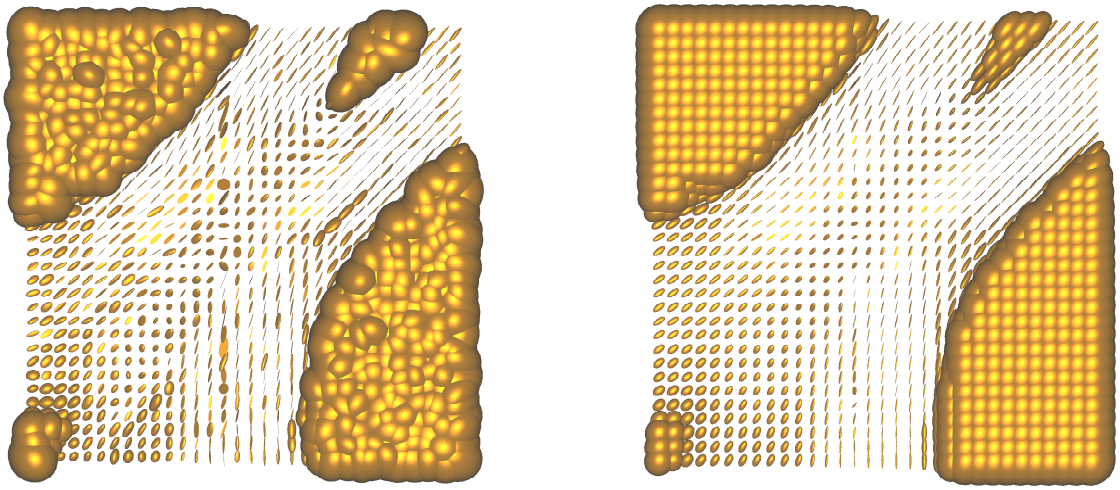


Figure 2.4: Denoising of a DT-MRI image.

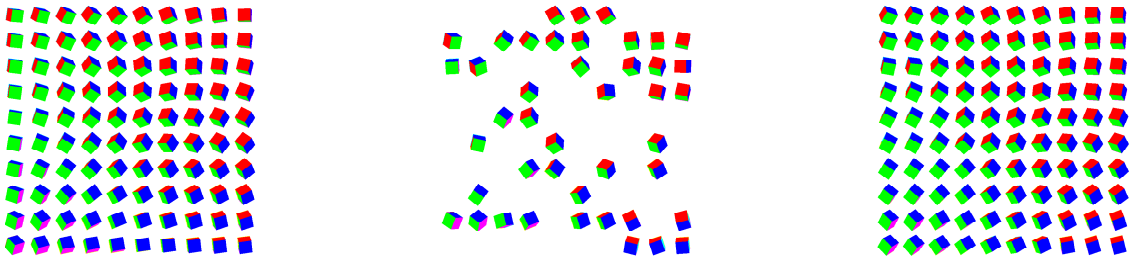
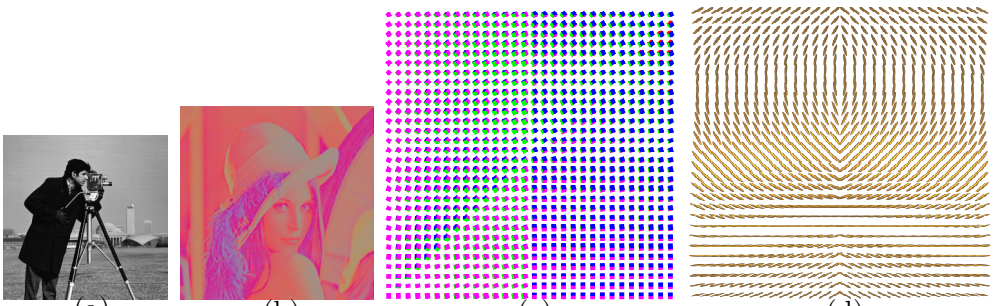


Figure 2.5: Denoising and inpainting of an $SO(3)$-valued image. From left to right: original, noisy and restored image.

2.6.3 Comparison to proximal point algorithm

We compared the IRM algorithm to the proximal point algorithm for the four images shown in Figure 2.6. Images (a) and (b) are well known in the image processing community. In image (c), for $(i,j) \in \{1,\dots,30\}^2$ the value of the pixel (i,j) is given by the rotation matrix with axis $(2(i-1),\ j-1,\ 0)$ if $i \geq 16$ and $(0,\ 2(i-1),\ 14.5)$ if $i \leq 15$ and angle $(i+j-2)/29$ if $i > j$ and $\pi/2 + (i-j)/29$ if $i \leq j$. In image (d), for $i \in \{1,\dots,15\}$ and $j \in \{1,\dots,30\}$ the value of the pixel (i,j) is RDR^T where $D = \mathrm{diag}((1.7,\ 0.3,\ 0.2))$ and R is the rotation matrix with axis $((i-15.5)/5,\ 3(j-15.5),\ 29)$ and angle $3\pi/4 + (i-15.5)/5 - 3(j-15.5)$. The value of pixel $(31-i,j)$ is constructed similar as the value of the pixel (i,j), with the only difference that the angle of rotation is $-3\pi/4 + (i-15.5)/5 + 3(j-15.5)$. For each image, we added noise with standard deviation 0.2. For the sphere- and $SO(3)$-valued image, we projected the result back on

the manifold. For the $SPD(3)$-valued image, we computed the matrix logarithm, added noise and then computed the matrix exponential of the result. We choose $\lambda = 0.2$ and the sequence $\mu_k = 3k^{-0.95}$ for all four experiments. We measured the computational time, the value of J and the the peak-to-signal noise ratio (PSNR) with respect to the original image for both algorithms. The code implemented is far from being optimal. There is, for example, no parallelization used. Both algorithms use however the same subroutines, which makes a comparison feasible. In Figure 2.7, we plotted the results.



(a) (b) (c) (d)

Figure 2.6: Test images:(a) Cameraman (256×256, $\mathbb{R}$-valued) (b) Color Part of Lena (361×361, S^2-valued) (c) Synthetic 30×30 $SO(3)$-valued image (d) Synthetic 30×30 $SPD(3)$-valued image

For all four images, IRM computes the minimizer of J faster than proximal point. However, the PSNR is not always smaller in the IRM case. The reason is that the PSNR is sometimes larger during the algorithm than at the end, when we are close to the TV minimizer. This, however, depends on the choice of $(\mu_k)_{k\in\mathbb{N}}$ and the image.

In [19] a C++ template library for the minimization of the TV-functional using IRM and proximal point was implemented and a similar experiment was conducted. The observation was that for manifolds for which the exponential and logarithm map are simple to compute (e.g. Euclidean data or the sphere) proximal point is faster whereas for manifolds for which the computation of the exponential and logarithm map is more expensive ($SO(n)$, $SPD(n)$ or the Grassmannian) IRM becomes faster. The explanation is that for IRM the majority of the computational time is used to solve the linear system of the Newton step which does not suffer from increasing complexity of exp and log, whereas for proximal point the majority of the computational time is used to compute exp and log.

If the IRM algorithm is applied to a very large image the amount of storage can be a limiting factor. To avoid this issue, it is recommended to subdivide the image into smaller subimages and apply the IRM algorithm independently to all these subimages. If desired, this process can of course also be parallelized.

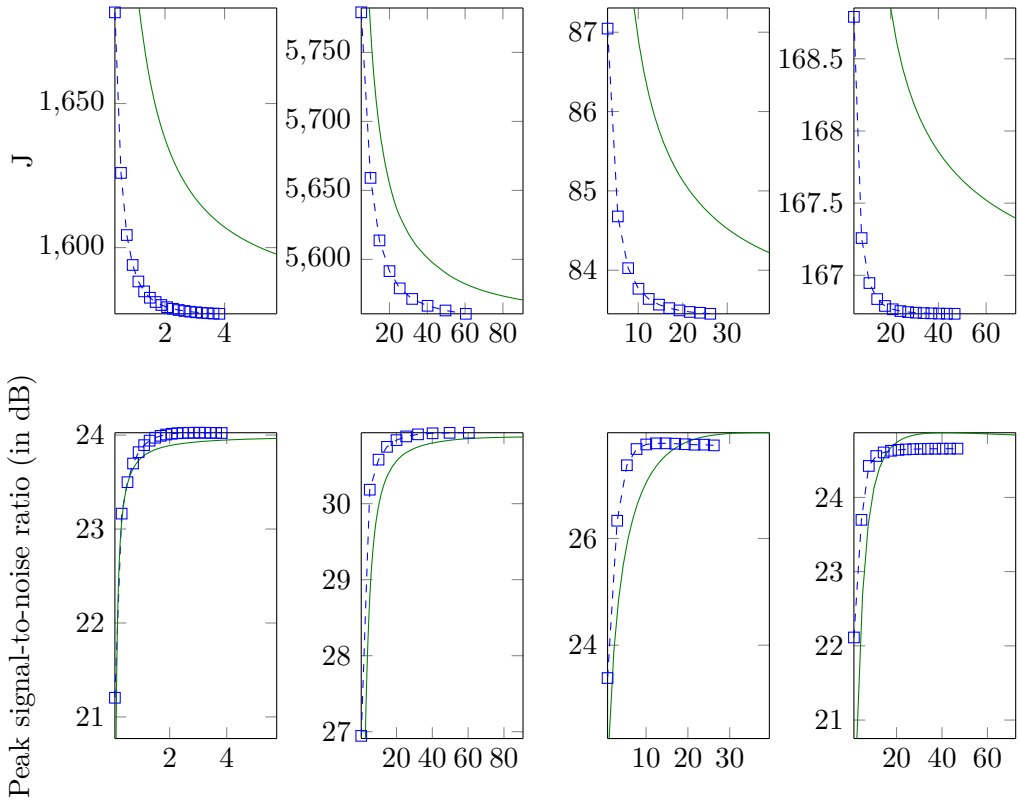


Figure 2.7: On top the value of the functional J and at the bottom the PSNR with dependence on the computational time (in second) is shown. From left to right: Results for images (a), (b), (c) and (d). The blue dashed lines with squares is IRM and the green line proximal point.

Much more experimentation would be needed to evaluate the IRM method relative to other techniques. What is demonstrated in this chapter is simply that IRM can be used for the tasks presented.

3 Approximation Error Estimates

In this chapter, we study approximations of manifold-valued functions defined on a domain $\Omega \subset \mathbb{R}^s$. For our construction, we will need basis functions $\phi_i \colon \Omega \to \mathbb{R}$ where i is an element of an index set I. With the Riemannian and the projection-based average introduced in Section 1.4, we can generate two geometric finite element spaces V_R and $V_{\mathcal{P}}$ as defined in (0.7), i.e.

$$V_R := \{v \colon \Omega \to M,\ v(x) := av_{Riem}\left((\phi_i(x))_{i\in I}, (p_i)_{i\in I}\right) | p_i \in M \text{ for all } i \in I\},$$

and

$$V_{\mathcal{P}} := \{v \colon \Omega \to M,\ v(x) := av_{\mathcal{P}}\left((\phi_i(x))_{i\in I}, (p_i)_{i\in I}\right) | p_i \in M \text{ for all } i \in I\}.$$

In this chapter, we study how well a function can be approximated by a function in V_R respectively $V_{\mathcal{P}}$. To this end we will construct approximation operators into these spaces. We will assume that for all $i \in I$ the basis function $\phi_i \colon \Omega \to \mathbb{R}$ is supported locally around a point $\xi_i \in \bar{\Omega}$, where $\bar{\Omega}$ denotes the closure of Ω. Then we can define natural approximation operators for continuous functions into V_R respectively $V_{\mathcal{P}}$.

Definition 3.0.1. Let $\Omega \subset \mathbb{R}^s$, M a Riemannian manifold. The approximation operator $Q_R \colon C(\bar{\Omega}, M) \to V_R$ respectively $Q_{\mathcal{P}} \colon C(\bar{\Omega}, M) \to V_{\mathcal{P}}$ corresponding to a set of basis functions $\Phi = (\phi_i)_{i\in I} \colon \Omega \to \mathbb{R}$ and nodes $\Xi = (\xi_i)_{i\in I} \subset \bar{\Omega}$ is defined by

$$Q_R u(x) := av_{Riem}((u(\xi_i))_{i\in I}, (\phi_i(x))_{i\in I}) \text{ for all } x \in \Omega, \tag{3.1}$$

respectively

$$Q_{\mathcal{P}} u(x) := av_{\mathcal{P}}((u(\xi_i))_{i\in I}, (\phi_i(x))_{i\in I}) \text{ for all } x \in \Omega. \tag{3.2}$$

Note that for $u \in C(\bar{\Omega}, M)$ and functions ϕ_i with sufficiently small support the functions $Q_R u$ and $Q_{\mathcal{P}} u$ are well-defined. We will also consider the linear analog $Q_{\mathbb{R}^n} \colon C(\bar{\Omega}, M) \to V_{\mathbb{R}^n} = \{\sum_{i\in I} \phi_i c_i \mid c_i \in \mathbb{R}^n\}$ of $Q_R u$ and $Q_{\mathcal{P}} u$ defined by

$$Q_{\mathbb{R}^n} u := \sum_{i\in I} \phi_i u(\xi_i). \tag{3.3}$$

Note that we have $Q_{\mathcal{P}} = \mathcal{P} \circ Q_{\mathbb{R}^n}$ where by abuse of notation $\mathcal{P}$ also denotes the operator $\mathcal{P} \colon C(\bar{\Omega}, \mathbb{R}^n) \to C(\bar{\Omega}, M)$ which applies the closest point projection $\mathcal{P}$ pointwise.

We briefly describe an example of $Q_{\mathcal{P}}$. Assume that $\Omega \subset \mathbb{R}^2$ is a polygonal domain. Consider a triangulation of Ω and denote its vertices by $(\xi_i)_{i\in I}$. For every $i \in I$ we

define ϕ_i as the piecewise (on every triangle) linear and globally continuous function with $\phi_i(\xi_j) = \delta_{ij}$. Let now $u\colon \Omega \to S^1 \subset \mathbb{R}^2$ be a map into the circle S^1. Consider Figure 3.1. Since the basis functions are piecewise linear the restriction of the function $Q_{\mathbb{R}^2}u = \sum_{i\in I} \phi_i u(\xi_i)$ to the triangle with vertices ξ_i, ξ_k and ξ_m (in red) is an affine map. Its image is the triangle with vertices $u(\xi_i), u(\xi_k)$ and $u(\xi_m)$ (in green). Let x be in the red triangle and $Q_{\mathbb{R}^2}u(x)$ be its image in the green triangle. To get the value $Q_{\mathcal{P}}u(x)$ we need to project $Q_{\mathbb{R}^2}u(x)$ onto the circle.

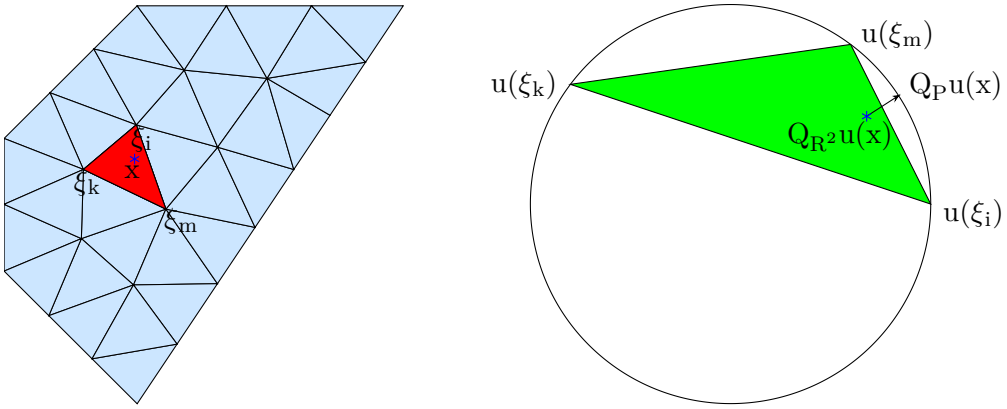


Figure 3.1: Geometric finite element function into S^1

The goal of this chapter is to estimate the error (measured in a Sobolev norm) between a function $u\colon \Omega \to M$ and its approximation $Q_R u \in V_R$ respectively $Q_{\mathcal{P}}u \in V_{\mathcal{P}}$ in terms of the mesh width

$$h := \sup_{x\in\Omega} \max_{\substack{i\in I \\ \phi_i(x)\neq 0}} |\xi_i - x|,$$

and norms of u.

A classical estimate (see e.g. [48, 9]) states that for $p \in [1,\infty]$, $m > \frac{s}{p}$ and $l \leq m$ we have under some assumptions (e.g. polynomial exactness, smoothness) on $(\phi_i)_{i\in I}$ and $(\xi_i)_{i\in I}$ that

$$|u - Q_{\mathbb{R}^n}u|_{W^{l,p}} \lesssim h^{m-l}|u|_{W^{m,p}} \text{ for all } u \in W^{m,p}(\Omega, \mathbb{R}^n), \tag{3.4}$$

In Section 3.3, respectively 3.4, we will prove that $Q_{\mathcal{P}}$ respectively Q_R satisfy a similar error estimate. The idea to prove an approximation error estimate for $Q_{\mathcal{P}} = \mathcal{P}Q_{\mathbb{R}^n}$ is

to first show that the operator $\mathcal{P}$ is locally Lipschitz continuous with respect to Sobolev norms $W^{l,p}$, i.e.

$$|\mathcal{P}v - \mathcal{P}w|_{W^{l,p}} \lesssim \|v - w\|_{W^{l,p}}. \tag{3.5}$$

The exact statement also including the dependence of the constant will be presented and proven in Section 3.2. Then we combine this estimate with (3.4) in Section 3.3 to get

$$|u - Q_{\mathcal{P}}u|_{W^{l,p}} = |\mathcal{P}u - \mathcal{P}Q_{\mathbb{R}^n}u|_{W^{l,p}} \lesssim h^{m-l}|u|_{W^{m,p}}.$$

More precisely, we can show that asymptotically (i.e. for h small enough) $Q_{\mathcal{P}}$ satisfies the same error estimate as $Q_{\mathbb{R}^n}$ does (3.4).

In Section 3.4, we will prove a similar error estimate for Q_R. However, we will not be able to prove the exact same estimate as for $Q_{\mathcal{P}}$. Instead, we estimate the error by h^{m-l} times a factor depending only on the derivatives of u. In Section 3.5, we will consider the case where $\Omega = \mathbb{R}$ and $(\phi_i)_{i\in I}$ are B-splines. Here the standard approximation operator $Q_{\mathbb{R}^n}$ does not yield optimal approximation order and another type of operator needs to be generalized.

3.1 Properties of projection-based finite element spaces

In this section we study some simple properties of projection-based finite elements. By definition every function $v \in V_{\mathcal{P}}$ is of the form

$$v(x) = \mathcal{P}\left(\sum_{i\in I} c_i\phi_i(x)\right) \text{ for all } x \in \Omega \text{ and } (c_i)_{i\in I} \subset M. \tag{3.6}$$

Hence each function $v \in V_{\mathcal{P}}$ is characterized by the values $(c_i)_{i\in I} \subset M$. In Section 3.1.1, we show that projection-based finite elements are conforming. In Section 3.1.2, we show that projection-based finite elements are equivariant under certain isometries of M.

3.1.1 Conformity of projection-based finite element spaces

By the chain rule we have

$$\frac{\partial}{\partial_{x_j}} v(x) = \mathcal{P}'\left(\sum_{i\in I} c_i\phi_i(x)\right)\left[\sum_{i\in I} c_i \frac{\partial}{\partial_{x_j}}\phi_i(x)\right] \tag{3.7}$$

for every $x \in \Omega$ such that $\mathcal{P}$ is differentiable at $\sum_{i\in I} c_i\phi_i(x) \in \mathbb{R}^n$ and ϕ_i is differentiable at $x \in \Omega$ for all $i \in I$. We can now prove that $V_{\mathcal{P}}$ is $W^{1,p}$ conforming.

Proposition 3.1.1. *Assume that the closest point projection $\mathcal{P}\colon U \subset \mathbb{R}^n \to M$ is differentiable with bounded derivative and $\Phi = (\phi_i)_{i\in I} \subset W^{1,p}$ for some $p \in [1,\infty]$. Then we have $V_{\mathcal{P}} \subset W^{1,p}$.*

Remark 3.1.2. *In [1] it was shown that if M is a closed C^2-manifold with radius of curvature bounded from below the closest point projection defined in Definition 1.2.2 is well-defined and differentiable with bounded derivative in a uniform neighborhood of M. Hence the first assumption of Lemma 3.1.1 is satisfied in this case.*

Proof. Let $v \in V_{\mathcal{P}}$ and $(c_i)_{i\in I} \subset M$ such that (3.6) holds. As functions in $W^{1,p}$ are differentiable a.e. the function $v \in V_{\mathcal{P}}$ is also differentiable a.e. and the weak derivative coincides with the classical derivative (given in (3.7)) a.e. (see for example Corollary 8.11 of [10]). Because I is finite we have for $j \in \{1,\dots,d\}$ and $p < \infty$

$$\left\|\frac{\partial}{\partial_{x_j}} v\right\|_{L^p}^p = \int_{x\in\Omega} \left|\mathcal{P}^{(1)}\left(\sum_{i\in I} c_i\phi_i(x)\right)\left[\sum_{i\in I} c_i \frac{\partial}{\partial_{x_j}}\phi_i(x)\right]\right|^p dx \lesssim |\mathcal{P}|_{C^1}^p \sum_{i\in I} |c_i|^p \left\|\frac{\partial}{\partial_{x_j}}\phi_i\right\|_{L^p}^p < \infty,$$

For the case $p = \infty$ we get for $j \in \{1,\dots,d\}$

$$\left\|\frac{\partial}{\partial_{x_j}} v\right\|_{L^\infty} = \sup_{x\in\Omega} \left|\mathcal{P}^{(1)}\left(\sum_{i\in I} c_i\phi_i(x)\right)\left[\sum_{i\in I} c_i \frac{\partial}{\partial_{x_j}}\phi_i(x)\right]\right| \lesssim |\mathcal{P}|_{C^1}^p \sum_{i\in I} |c_i| \sup_{x\in\Omega} \left\|\frac{\partial}{\partial_{x_j}}\phi_i\right\|_{L^\infty} < \infty. \quad \square$$

We now study higher derivatives of v. The l-th derivative $\mathcal{P}^{(l)}(x)$ of $\mathcal{P}$ at $x \in \mathbb{R}^n$ is a multilinearform which we denote by $\mathcal{P}^{(l)}[\cdot,\dots,\cdot]$. A derivative $D^{\vec{a}}v(x)$ at $x \in \Omega$ is a sum of terms of the form

$$\mathcal{P}^{(l)}(Q_{\mathbb{R}^n}v(x))[D^{\vec{a_1}}Q_{\mathbb{R}^n}v(x),\dots,D^{\vec{a_l}}Q_{\mathbb{R}^n}v(x)]$$

where $l \le |\vec{a}|_1$, $\vec{a_i} \in \mathbb{N}^d\backslash\{(0,\dots,0)\}$ with $\vec{a_1} + \cdots + \vec{a_l} = \vec{a}$. Estimating the L^p-norm of these expressions is not as simple as before. It is not even true that for smooth function $G\colon \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$, $n_1, n_2 \in \mathbb{N}$ and $v \in W^{l,p}(\Omega,\mathbb{R}^{n_1})$ we have $Gv \in W^{l,p}(\Omega,\mathbb{R}^{n_2})$ as the following example shows.

Example 3.1.3. *Let $s = 3$, $n_1 = n_2 = 1$, $\Omega = B_1(0) := \{x \in \mathbb{R}^s \mid |x| < 1\}$, $v(x) := |x|^{-0.5}$ and $G(x) := \sin(x)$. We have*

$$\int_{B_1(0)} \left|\frac{d^2}{dx_i dx_j} v(x)\right| dx = \int_{B_1(0)} \left|\frac{5}{4}|x|^{-\frac{9}{2}}x_i x_j - \frac{1}{2}|x|^{-\frac{5}{2}}\delta_{ij}\right| dx \lesssim \int_{B_1(0)} |x|^{-\frac{5}{2}} dx \lesssim \int_0^1 r^{-\frac{1}{2}} dr \lesssim 1.$$

Hence $v \in W^{2,1}$. However

$$\begin{aligned}
\int_{B_1(0)} |\Delta G(v(x))| dx &= \int_{B_1(0)} \left|\left(\frac{5}{4} - \frac{3}{2}\right)\cos(|x|^{-0.5})|x|^{-\frac{5}{2}} - \frac{1}{4}\sin(|x|^{-0.5})|x|^{-3}\right| dx \\
&\sim \int_0^1 \left|\cos(r^{-0.5})r^{-\frac{5}{2}} + \sin(r^{-0.5})r^{-3}\right| r^2 dr \\
&\sim \int_1^\infty \left|\cos(t)t^{-2} + \sin(t)t^{-1}\right| dt \\
&\ge \int_1^\infty \left|\frac{\sin(t)}{t}\right| dt - \int_1^\infty t^{-2} dt \\
&= \infty.
\end{aligned}$$

Hence $Gv \notin W^{2,1}$.

However, if we assume that our basis functions $(\phi_i)_{i\in I}$ are bounded (i.e. in L^∞) we can show that $V_\mathcal{P} \subset W^{l,p}$.

Proposition 3.1.4. *Assume that the closest point projection $\mathcal{P}$ is l times differentiable with bounded derivatives and $\Phi = (\phi_i)_{i\in I} \subset W^{l,p} \cap L^\infty$. Then we have $V_\mathcal{P} \subset W^{1,p} \cap L^\infty$.*

Proof. Let $v \in V_\mathcal{P}$ and $(c_i)_{i\in I} \subset M$ such that (3.6) holds. Then there exists $w \in C(\Omega, M)$ with $v = Q_M w = \mathcal{P} Q_{\mathbb{R}^n} w$. We have $Q_{\mathbb{R}^n} w = \sum_{i\in I} c_i \phi_i \in W^{l,p} \cap L^\infty$. By the Corollary of the first Theorem in Section 5.2.5 of [45] we have that $v = \mathcal{P} Q_{\mathbb{R}^n} w \in W^{l,p} \cap L^\infty$. □

3.1.2 Preservation of isometries

In this section, we study under which circumstances an isometry $T\colon M \to M$ commutes with the projection-based interpolation operator Q_M. The projection-based finite element space $V_\mathcal{P}$ defined in (0.7) is then called equivariant under the isometry T. In mechanics, this leads to the desirable property that discretizations of objective problems are again objective.

We will need our isometries to be extendable as defined below.

Definition 3.1.5. An isometry $T\colon M \to M$ (with respect to the geodesic distance) is called extendable if there exists an isometry $\tilde{T}\colon \mathbb{R}^n \to \mathbb{R}^n$ (with respect to the Euclidean distance) with $\tilde{T}(p) = T(p)$ for all $x \in M$.

Examples for extendable isometries T are orthogonal transformations for the sphere and multiplication with special orthogonal matrices for $SO(n)$. We can now prove our main theorem of this section

Theorem 3.1.6. *Let $M \subset \mathbb{R}^n$ be a Riemannian submanifold, $\mathcal{P}$ the closest-point projection defined in Definition 1.2.2, $T\colon M \to M$ an extendable isometry and $\Phi = (\phi_i)_{i\in I}$ a partition of unity. Then T commutes with Q_M.*

Proof. As an isometry maps closest distances to closest distances, $\tilde{T}$ commutes with $\mathcal{P}$. Therefore $\tilde{T}$ commutes with $\mathcal{P}$. By the Mazur–Ulam theorem [33] there exists a linear map $A\colon \mathbb{R}^n \to \mathbb{R}^n$ with $\tilde{T}(q) = A(q) + \tilde{T}(0)$ for all $q \in \mathbb{R}^n$. We now have for any $v \in C(\Omega, M)$ and $x \in \Omega$

$$\begin{aligned}\tilde{T} Q_{\mathbb{R}^n} v(x) &= \tilde{T} \sum_{i\in I} v(\xi_i)\phi_i(x) = A \sum_{i\in I} v(\xi_i)\phi_i(x) + \tilde{T}(0) = \sum_{i\in I} Av(\xi_i)\phi_i(x) + \tilde{T}(0) \sum_{i\in I} \phi_i(x) \\ &= \sum_{i\in I} (Av(\xi_i) + \tilde{T}(0))\phi_i(x) = \sum_{i\in I} \tilde{T} v(\xi_i)\phi_i(x) = Q_{\mathbb{R}^n} \tilde{T} v(x).\end{aligned}$$

Hence $\tilde{T}$ commutes with $Q_{\mathbb{R}^n}$. We now have

$$T \circ Q_M = \tilde{T} \circ Q_M = \tilde{T} \circ \mathcal{P} \circ Q_{\mathbb{R}^n} = \mathcal{P} \circ \tilde{T} \circ Q_{\mathbb{R}^n} = \mathcal{P} \circ Q_{\mathbb{R}^n} \circ \tilde{T} = Q_M \circ \tilde{T} = Q_M \circ T. \quad \square$$

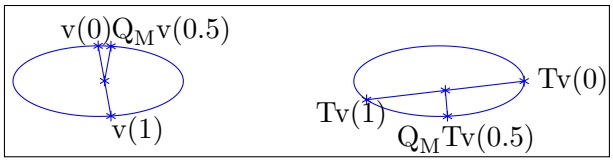


Figure 3.2: An isometry not preserved by Q_M

However, in general Q_M does not commute with S_T as can be seen by the example illustrated in Figure 3.2. In this example M is an ellipse in $\mathbb{R}^2$, $\mathcal{P}$ the closest point projection and T the isometry which moves every point clockwise a quarter of the total length of the ellipse. Suppose that $Q_{\mathbb{R}^2}$ (and therefore also Q_M) is exact at 0 and 1 and $Q_{\mathbb{R}^2} v(0.5) = (v(0) + v(1))/2$. Then we can see that $Q_M v(0.5) = \mathcal{P}((v(0) + v(1))/2)$ is close to $v(0)$. However, $Q_M(T(v(0.5)))$ is roughly in the middle of $Tv(0)$ and $Tv(1)$ and therefore not equal to $T(Q_M(v(0.5)))$.

3.2 Lipschitz continuity of Composition Operators

In this section, our goal is to show local Lipschitz continuity of $\mathcal{P}$ i.e. (3.5). In fact, we will show more generally that (3.5) also holds if we replace the closest point projection $\mathcal{P}$ by any sufficiently regular function $G\colon U \subset \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$. The bound of the error will also include the Lipschitz constants of the derivatives of G.

Definition 3.2.1. Let $n_1, n_2, l \in \mathbb{N}$, $G\colon U \subset \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ be l-times differentiable with bounded Lipschitz continuous derivatives. We define

$$\operatorname{Lip}\left(G^{(l)}\right) := \sup_{\substack{q,r \in U \\ q \neq r}} \frac{\left\| G^{(l)}(q) - G^{(l)}(r) \right\|}{|q - r|}.$$

The following lemma treats the special case $l = 0$. It shows that G is globally Lipschitz continuous with respect to the L^p-norm and the Lipschitz constant can be chosen equal to $\operatorname{Lip}(G)$.

Lemma 3.2.2. *Let $\Omega \subset \mathbb{R}^s$ be an open and bounded Lipschitz domain, $G\colon U \subset \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ be Lipschitz continuous and $p \in [1, \infty]$. Then we have*

$$\|Gv - Gw\|_{L^p} \leq \operatorname{Lip}(G) \|v - w\|_{L^p}$$

for all $v, w \in L^p(\Omega, U)$.

Proof. By the Definition of $\operatorname{Lip}(G)$ we have

$$\begin{aligned} \|Gv - Gw\|_{L^p} &= \| |Gv - Gw|_2 \|_{L^p} \leq \| \operatorname{Lip}(G) |v - w|_2 \|_{L^p} = \operatorname{Lip}(G) \| |v - w|_2 \|_{L^p} \\ &= \operatorname{Lip}(G) \|v - w\|_{L^p}. \quad \square \end{aligned}$$

Next we treat the case $l = 1$. The idea is to rewrite the derivatives at $x \in \Omega$ as a telescopic sum, i.e.

$$\begin{aligned}\frac{\partial}{\partial_{x_j}}(Gv(x) - Gw(x)) &= G^{(1)}(v(x))\left[\frac{\partial}{\partial_{x_j}}v(x)\right] - G^{(1)}(w(x))\left[\frac{\partial}{\partial_{x_j}}w(x)\right] \\ &= \left(G^{(1)}(v(x)) - G^{(1)}(w(x))\right)\left[\frac{\partial}{\partial_{x_j}}v(x)\right] \\ &\quad + G^{(1)}(w(x))\left[\frac{\partial}{\partial_{x_j}}(v(x) - w(x))\right].\end{aligned}$$

Taking the L^p-norm we can estimate the first term using the Hölder inequality and the Sobolev embedding theorem by $\mathrm{Lip}(G^{(1)})\|v - w\|_{L^r}\|v\|_{W^{m,p}}$. The second term can be estimated by the term $|G|_{C^1}\|v - w\|_{W^{1,p}}$. Furthermore, $r \in [1, \infty]$ can be chosen such that $W^{1,p} \subset L^r$.

Lemma 3.2.3. *Let $\Omega \subset \mathbb{R}^s$ be an open and bounded Lipschitz domain, $G\colon U \subset \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ a differentiable function with bounded and Lipschitz continous derivative, $p \in [1, \infty]$ and $m > \frac{s}{p}$. Then there exists $r \in [p, \infty]$ with $\frac{1}{r} > \frac{1}{p} - \frac{1}{s}$ such that we have*

$$|Gv - Gw|_{W^{1,p}} \leq |G|_{C^1}|v - w|_{W^{1,p}} + C\,\mathrm{Lip}(G^{(1)})\|v - w\|_{L^r}\|v\|_{W^{m,p}}$$

for all $v \in W^{m,p}(\Omega, U)$ and $w \in W^{1,p}(\Omega, U)$ where $C > 0$ is a constant depending only on m, s, p and Ω.

Proof. By the triangle inequality it is enough to prove that

$$\left\|\frac{\partial}{\partial_{x_j}}(Gv - Gw)\right\|_{L^p} \leq |G|_{C^1}\left\|\frac{\partial}{\partial_{x_j}}(v - w)\right\|_{L^p} + C\,\mathrm{Lip}(G^{(1)})\|v - w\|_{L^s}\left\|\frac{\partial}{\partial_{x_j}}v\right\|_{W^{m-1,p}}$$

for all $j \in \{1, \ldots, s\}$. Choose

$$0 < \epsilon \leq \frac{1}{s}\min\left(1, m - \frac{s}{p}\right), \quad \frac{1}{r} := \max\left(0, \frac{1}{p} - \frac{1}{s} + \epsilon\right), \quad \text{and} \quad \frac{1}{t} := \max\left(0, \frac{1}{p} - \frac{m-1}{s}\right).$$

In order to be able to apply Hölder we prove the inequality

$$\frac{1}{r} + \frac{1}{t} \leq \frac{1}{p}, \tag{3.8}$$

which follows from

$$\frac{1}{r} + \frac{1}{t} \in \left\{0, \frac{1}{p} - \frac{1}{s} + \epsilon, \frac{1}{p} - \frac{m-1}{s}, \frac{2}{p} - \frac{m}{s} + \epsilon\right\},$$

$0 \leq \frac{1}{p}$, $\frac{1}{p} - \frac{1}{s} + \epsilon \leq \frac{1}{p}$, $\frac{1}{p} - \frac{m-1}{s} \leq \frac{1}{p}$ and $\frac{2}{p} - \frac{m}{s} + \epsilon = \frac{1}{p} - \frac{1}{s}\left(m - \frac{s}{p}\right) + \epsilon \leq \frac{1}{p}$. For $f\colon \Omega \to U$ and $g\colon \Omega \to \mathbb{R}^{n_1}$ we define $G^{(1)}(f)[g]\colon \Omega \to \mathbb{R}^{n_2}$ by

$$G^{(1)}(f)[g](x) := G^{(1)}(f(x))[g(x)].$$

Then we have by the chain rule, the triangle inequality, Hölder's inequality and the Sobolev embedding theorem that

$$\begin{aligned}
\left\|\frac{\partial}{\partial_{x_j}}(Gv-Gw)\right\|_{L^p} &= \left\|G^{(1)}(v)\left[\frac{\partial}{\partial_{x_j}}v\right]-G^{(1)}(w)\left[\frac{\partial}{\partial_{x_j}}w\right]\right\|_{L^p} \\
&= \left\|\left(G^{(1)}(v)-G^{(1)}(w)\right)\left[\frac{\partial}{\partial_{x_j}}v\right]+G^{(1)}(w)\left[\frac{\partial}{\partial_{x_j}}(v-w)\right]\right\|_{L^p} \\
&\leq \operatorname{Lip}(G^{(1)})\left\||v-w|\cdot\left|\frac{\partial}{\partial_{x_j}}v\right|\right\|_{L^p}+|G|_{C^1}\left\|\frac{\partial}{\partial_{x_j}}(v-w)\right\|_{L^p} \\
&\lesssim C\operatorname{Lip}(G^{(1)})\|v-w\|_{L^r}\left\|\frac{\partial}{\partial_{x_j}}v\right\|_{L^t}+|G|_{C^1}\left\|\frac{\partial}{\partial_{x_j}}(v-w)\right\|_{L^p} \\
&\lesssim |G|_{C^1}\left\|\frac{\partial}{\partial_{x_j}}(v-w)\right\|_{L^p}+C\operatorname{Lip}(G^{(1)})\|v-w\|_{L^r}\|v\|_{W^{m,p}}. \quad \square
\end{aligned}$$

Another proof for Lemma 3.2.3 can be found in [32]. Estimating $|Gv-Gw|_{W^{l,p}}$ for $l\geq 2$ is more difficult. The problem when trying to prove it in a similar fashion as we did in Lemma 3.2.3 is that products with partial derivatives of w will occur for which we do not want to assume the same smoothness as for v since in Section 3.3 we will apply this Lemma with $w=Q_{\mathbb{R}^n}v$. However, if we assume a slightly higher smoothness than $W^{l,p}$ of w we can prove a similar statement as Lemma 3.2.3 also for $l\geq 2$. It will seem like a very lucky coincidence that the following estimates work in way that for the second term on the right hand side we can use a weaker norm of $v-w$ than $W^{l,p}$. It is likely that there exists a deeper reason for the statement to hold true which however still needs to be found.

Lemma 3.2.4. *Let $\Omega\subset\mathbb{R}^s$ be an open and bounded Lipschitz domain, $m>\frac{s}{p}$, $2\leq l\leq m$, $G\colon U\subset\mathbb{R}^{n_1}\to\mathbb{R}^{n_2}$ be l times differentiable with bounded and Lipschitz continuous derivatives, $p\in[1,\infty]$ and $q\in[p,\infty]$ with $q>\frac{s}{l}$. Then there exists $r\in[p,\infty]$ with $\frac{1}{r}>\frac{1}{p}-\frac{1}{s}$ such that we have*

$$|Gv-Gw|_{W^{l,p}}\leq |G|_{C^1}|v-w|_{W^{l,p}}+CL(G)B\left(\|v\|_{W^{m,p}},\|w\|_{W^{l,q}}\right)\|v-w\|_{W^{l-1,r}}$$

for all $v\in W^{m,p}(\Omega,U)$ and $w\in W^{l,q}(\Omega,U)$ where C is a constant depending only on m,p,q,r and Ω,

$$L(G):=\max_{k\in\{1,\dots,l\}}\operatorname{Lip}\left(G^{(k)}\right)\quad\text{and}\quad B(x,y):=(1+x)^l+(1+x+y)^{l-1}. \tag{3.9}$$

Proof. Again by the triangle inequality it is enough to prove that

$$\left\|D^{\vec{a}}(Gv-Gw)\right\|_{L^p}\leq |G|_{C^1}\left\|D^{\vec{a}}(v-w)\right\|_{L^p}+CL(G)B\left(\|v\|_{W^{m,p}},\|w\|_{W^{l,q}}\right)\|v-w\|_{W^{l-1,r}} \tag{3.10}$$

for all $\vec{a} \in \mathbb{N}^s$ with $|\vec{a}|_1 = l$ where $D^{\vec{a}}$ was defined in (0.9). The derivative

$$D^{\vec{a}}(G(v(x)) - G(w(x)))$$

can be written as a sum of terms of the form

$$G^{(k)}(v(x)) \left[D^{\vec{a_1}}v(x), ..., D^{\vec{a_k}}v(x)\right] - G^{(k)}(w(x)) \left[D^{\vec{a_1}}w(x), ..., D^{\vec{a_k}}w(x)\right] \tag{3.11}$$

where $k \leq l$, $\vec{a_1}, \ldots, \vec{a_k} \in \mathbb{N}^s \backslash \{(0, \ldots, 0)\}$ and $\vec{a_1} + \cdots + \vec{a_k} = \vec{a}$. For $k = 1$ the L^p-norm of (3.11) can be estimated by the first term on the right hand side of (3.10). Assume now $k > 1$. Expression (3.11) can be written as a sum of terms of the form

$$(G^{(k)}(v(x)) - G^{(k)}(w(x))) \left[D^{\vec{a_1}}v(x), ..., D^{\vec{a_k}}v(x)\right] \tag{3.12}$$

and

$$G^{(k)}(w(x)) \left[D^{\vec{a_1}}v(x), ..., D^{\vec{a_{i-1}}}v(x), D^{\vec{a_i}}(v(x) - w(x)), D^{\vec{a_{i+1}}}w(x), ..., D^{\vec{a_k}}w(x)\right], \tag{3.13}$$

The norm of the term (3.12) respectively (3.13) can be estimated by

$$\text{Lip}\left(G^{(k)}\right) |v(x) - w(x)| \prod_{j=1}^{k} \left|D^{\vec{a_j}}v(x)\right| \tag{3.14}$$

respectively

$$\text{Lip}\left(G^{(k-1)}\right) \prod_{j=1}^{i-1} \left|D^{\vec{a_j}}v(x)\right| \left|D^{\vec{a_i}}(v(x) - w(x))\right| \prod_{j=i+1}^{k} \left|D^{\vec{a_j}}w(x)\right|, \tag{3.15}$$

where we used $|G|_{C^k} \leq \text{Lip}\left(G^{(k-1)}\right)$. By Lemma 3.2.5 there exists $r \in [p, \infty]$ with $\frac{1}{r} > \frac{1}{p} - \frac{1}{s}$ such that the L^p-norm of Expression (3.14) respectively (3.15) (without the Lipschitz constant) can be estimated by

$$\|v - w\|_{W^{l-1,r}} \prod_{i=1}^{k} \|v\|_{W^{m,p}} \lesssim \|v - w\|_{W^{l-1,r}} (1 + \|v\|_{W^{m,p}})^l$$

respectively

$$\|v - w\|_{W^{l-1,r}} \prod_{j=1}^{i-1} \|v\|_{W^{m,p}} \prod_{j=i+1}^{k} \|w\|_{W^{l,q}} \lesssim \|v - w\|_{W^{l-1,r}} \left(1 + \|v\|_{W^{m,p}} + \|w\|_{W^{l,q}}\right)^{l-1}.$$

□

In Lemma 3.2.4 we needed to bound the norm of a product of derivatives of v, w and $v - w$. The basic idea is to use Hölder's inequality and then the Sobolev embedding theorem. While we can assume $v \in W^{m,p}$ the function w is in general only l-times weakly differentiable. Furthermore, we want only the $W^{l,p}$-norm of the difference $v - w$ to appear on the right hand side. It turns out that for such an estimate to hold we need to assume that $w \in W^{l,q}$ for some $q > \frac{s}{l}$ and $q \geq p$. Note that by the Sobolev embedding theorem this implies that w is continuous.

Lemma 3.2.5. *Let $k > 0$ and for all $i \in \{0, \dots, k\}$ let $a_i, m_i,\ p_i \in [1, \infty]$. Assume that $m_i > \frac{s}{p_i}$, $p_i \geq p_0$, $l := \sum_{j=0}^{k} a_j \leq m_i$ for all $i \in \{0, \dots, k\}$ and $a_0 < l$. Then there exists $r \in [p_0, \infty]$ with $\frac{1}{r} > \frac{1}{p_0} - \frac{1}{s}$ such that*

$$\left\| \prod_{i=0}^{k} v_i \right\|_{L^{p_0}} \lesssim \|v_0\|_{W^{l-1-a_0,r}} \prod_{i=1}^{k} \|v_i\|_{W^{m_i-a_i,p_i}}$$

for all $v_i \in W^{m_i-a_i,p_i}$, $i \in \{0, \dots, k\}$.

Proof. Let

$$\begin{aligned}
0 < \epsilon &\leq \frac{1}{s} \min\left(1, \min_{i \in \{0,\dots,k\}} m_i - \frac{s}{p_i}\right), \\
\frac{1}{r} &:= \max\left(\frac{1}{p_0} - \frac{1}{s} + \epsilon, 0\right), \\
\frac{1}{t_0} &:= \max\left(\frac{1}{r} - \frac{l-1-a_0}{s}, 0\right), \\
\frac{1}{t_i} &:= \max\left(\frac{1}{p_i} - \frac{m_i - a_i}{s}, 0\right) \text{ for all } i \in \{1, \dots, k\} \text{ and} \\
A &:= \{i \in \{1, \dots, k\} \mid t_i < \infty\}.
\end{aligned}$$

In order to be able to apply Hölder's inequality we prove that

$$\sum_{i=0}^{k} \frac{1}{t_i} \leq \frac{1}{p_0}.$$

If $\frac{1}{t_0} > 0$ and $|A| = 0$ we have

$$\sum_{i=0}^{k} \frac{1}{t_i} = \frac{1}{t_0} = \frac{1}{s} - \frac{l-1-a_0}{s} \leq \frac{1}{r} \leq \frac{1}{p_0}.$$

If $\frac{1}{t_0} > 0$ and $|A| \geq 1$ we have $\frac{1}{r} > 0$ and hence $\frac{1}{r} = \frac{1}{p_0} - \frac{1}{s} + \epsilon$. Therefore, $\frac{1}{t_0} = \frac{1}{r} - \frac{l-1-a_0}{s} = \frac{1}{p_0} - \frac{l-a_0}{s} + \epsilon$ and

$$\begin{aligned}
\sum_{i=0}^{k} \frac{1}{t_i} &\leq \frac{1}{t_0} + \sum_{i \in A} \frac{1}{t_i} \\
&= \frac{1}{p_0} + \epsilon - \frac{l - a_0 - \sum_{i \in A} a_i}{s} - \frac{1}{s} \sum_{i \in A} \left(m_i - \frac{s}{p_i}\right) \\
&\leq \frac{1}{p_0}.
\end{aligned}$$

If $\frac{1}{t_0} = 0$ and $|A| = 0$ we get

$$\sum_{i=0}^{k} \frac{1}{t_i} = 0 \leq \frac{1}{p_0}.$$

If $\frac{1}{t_0} = 0$ and $|A| \geq 1$ we choose $j \in A$ and obtain

$$\sum_{i=0}^{k} \frac{1}{t_i} = \frac{1}{p_j} - \frac{m_j - \sum_{i\in A} a_i}{s} - \frac{1}{s} \sum_{i \in A\setminus\{j\}} \left(m_i - \frac{s}{p_i} \right) \leq \frac{1}{p_j} \leq \frac{1}{p_0}.$$

Hence, by Hölder's inequality and the Sobolev embedding theorem we get in any case

$$\left\| \prod_{i=0}^{k} v_i \right\|_{L^{p_0}} \leq \prod_{i=0}^{k} \|v_i\|_{L^{t_i}} \lesssim \|v_0\|_{W^{l-1-a_0,r}} \prod_{i=1}^{k} \|v_i\|_{W^{m_i-a_i,p_i}}. \qquad \square$$

3.3 Error estimates for the approximation operator $Q_{\mathcal{P}}$

In this section, we estimate the approximation error $u - Q_{\mathcal{P}}u$. By Lemma 3.2.2 we can estimate the L^p-norm as stated in the following theorem.

Theorem 3.3.1. *Assume that* (3.4) *holds with* $l = 0$. *Let* $n \in \mathbb{N}$, $M \subset \mathbb{R}^n$ *be an embedded submanifold and* $\mathcal{P}\colon U \subset \mathbb{R}^n \to M$ *be a Lipschitz continuous projection onto* M. *Then we have*

$$\|u - Q_{\mathcal{P}}u\|_{L^p} \lesssim h^m \operatorname{Lip}(\mathcal{P}) |u|_{W^{m,p}} \text{ for all } u \in W^{m,p}(\Omega, M),$$

with the implicit constant of Inequality (3.4).

For $l \geq 1$ we will need a generalized version of Inequality (3.4) given by

$$|u - Q_{\mathbb{R}^n}u|_{W^{l,q}} \lesssim h^{m-l-s\left(\frac{1}{p}-\frac{1}{q}\right)} |u|_{W^{m,p}} \quad \text{for all } u \in W^{m,p}(\Omega, \mathbb{R}^n). \tag{3.16}$$

This follows from (3.4) using the Gagliardo-Nirenberg inequality [41]. The approximation error estimate for $l \geq 1$ is stated below.

Theorem 3.3.2. *Let* $l \in \mathbb{N}_{>0}$. *If* $l = 1$ *assume that* (3.4) *holds. If* $l \geq 2$ *holds assume that for some* $q \in [p, \infty]$ *with* $q > \frac{s}{l}$ *we have* $\Phi = (\phi_i)_{i\in I} \subset W^{l,q}$ *and Inequality* (3.16). *Let* $n \in \mathbb{N}$, $M \subset \mathbb{R}^n$ *an embedded submanifold,* $U \subset \mathbb{R}^n$ *a neighborhood of* M *and* $\mathcal{P}\colon U \to M$ *a projection onto* M. *Assume that* $\mathcal{P}$ *is* l*-times differentiable with bounded and Lipschitz continuous derivatives. Then there exists a constant* $\alpha > 0$ *such that*

$$|u - Q_{\mathcal{P}}u|_{W^{l,p}} \lesssim h^{m-l} |u|_{W^{m,p}} \left(|\mathcal{P}|_{C^1} + CL(\mathcal{P}) h^{\alpha} (1 + \|u\|_{W^{m,p}})^l \right),$$

for all $u \in W^{m,p}(\Omega, M)$ *with the implicit constant of Inequality* (3.4), $L(\mathcal{P})$ *as defined in* (3.9), *and* C *a constant depending only on* m, p, Ω *and the implicit constant of* (3.16).

Remark 3.3.3. *For* $l \geq 2$ *the assumption on the basis functions* $\Phi = (\phi_i)_{i\in I}$ *in Theorem 3.3.2 is stronger than the assumptions required to get* (3.4). *However, the assumption* $\Phi = (\phi_i)_{i\in I} \subset W^{l,p}$ *would not be strong enough to guarantee that* $Q_{\mathcal{P}}u \in W^{l,p}$. *The assumption implies that the basis functions* $\Phi = (\phi_i)_{i\in I}$ *are continuous.*

Proof of Theorem 3.3.2. If $l \geq 2$ we can assume without loss of generality that $l - \frac{s}{q} \leq m - \frac{s}{p}$ so that we have the embedding $W^{m,p} \subseteq W^{l,q}$. Then by (3.4) we have

$$\|Q_{\mathbb{R}^n} u\|_{W^{l,q}} \lesssim \|u\|_{W^{l,q}} \lesssim \|u\|_{W^{m,p}}.$$

Hence, by Lemma 3.2.4 (respectively Lemma 3.2.3) there exists $r \in [p, \infty]$ with $\frac{1}{r} > \frac{1}{p} - \frac{1}{s}$ such that

$$\begin{aligned} |u - Q_M u|_{W^{l,p}} &= |\mathcal{P}u - \mathcal{P}Q_{\mathbb{R}^n} u|_{W^{l,p}} \\ &\lesssim |\mathcal{P}|_{C^1} \, |u - Q_{\mathbb{R}^n} u|_{W^{l,p}} + L(\mathcal{P})(1 + \|u\|_{W^{m,p}})^l \, \|u - Q_{\mathbb{R}^n} u\|_{W^{l-1,r}} \, . \end{aligned}$$

By Theorem 3.5.5 and Corollary 3.16 we get

$$\begin{aligned} &|\mathcal{P}|_{C^1} \, |u - Q_{\mathbb{R}^n} u|_{W^{l,p}} + L(\mathcal{P})(1 + \|u\|_{W^{m,p}})^l \, \|u - Q_{\mathbb{R}^n} u\|_{W^{l-1,r}} \\ &\lesssim |\mathcal{P}|_{C^1} h^{m-l} |u|_{W^{m,p}} + L(\mathcal{P})(1 + \|u\|_{W^{m,p}})^l h^{m-(l-1)-s\left(\frac{1}{p}-\frac{1}{r}\right)} |u|_{W^{m,p}} \\ &\lesssim h^{m-l} |u|_{W^{m,p}} \left(|\mathcal{P}|_{C^1} + L(\mathcal{P}) h^{s\left(\frac{1}{r}-\left(\frac{1}{p}-\frac{1}{s}\right)\right)} (1 + \|u\|_{W^{m,p}})^l \right). \end{aligned}$$

□

Note that the implicit constants of our estimates are all independent of M. The only dependence of the inequalities on M are the factors $|\mathcal{P}|_{C^1}$ and $L(\mathcal{P})$. However, since $L(\mathcal{P})$ is multiplied with h^α it becomes irrelevant for $h \to 0$. By Proposition 1.2.4 we have

$$\begin{aligned} |\mathcal{P}|_{C^1} &= \sup_{p \in U} \left\| \mathcal{P}^{(1)}(p) \right\| \\ &\leq \sup_{p \in U} \left\| \mathcal{P}^{(1)}(\mathcal{P}(p)) \right\| + Lip\left(\mathcal{P}^{(1)}\right) |p - \mathcal{P}(p)| \\ &= 1 + Lip\left(\mathcal{P}^{(1)}\right) \sup_{p \in U} |p - \mathcal{P}(p)|. \end{aligned}$$

This shows that $|\mathcal{P}|_{C^1}$ is approximately one in a neighborhood of M. Together with Theorem 3.3.2 this yields that asymptotically the interpolation operator corresponding to the closest point projection satisfies the same error estimate as its linear analog.

Theorem 3.3.4. *Consider the same setting as in Theorem 3.3.2 with $\mathcal{P}$ being the closest point projection from Definition 1.2.2. Let $C > 0$ be larger than the implicit constant of Inequality* (3.4) *and $u \in W^{m,p}(\Omega, M)$. Then for h small enough we have*

$$|u - Q_{\mathcal{P}} u|_{W^{l,p}} \leq C h^{m-l} |u|_{W^{m,p}}.$$

Proof. This follows from Theorem 3.3.2 and (3.17). □

3.4 Error estimates for the approximation operator Q_R

In this section, we estimate the error between $Q_R u$ and u. By (1.14) we have

$$\sum_{i\in I} \phi_i(x)\log(Q_R u(x), u(\xi_i)) = 0 \tag{3.17}$$

The key idea to bound the approximation error of Q_R is to notice that the left hand side of the balance law (3.17) can be interpreted as the approximation of the function $y \mapsto \log(Q_R u(x), u(y))$ at $y = x$. There is also the alternate approach by directly estimating the approximation error of Q_R using Taylor expansion [24, 23].

In order to be able to work with classical derivatives we use that by Nash's isometric embedding theorem [39] we can assume that M is a Riemannian submanifold of $\mathbb{R}^n$. Furthermore we extend the logarithm $\log\colon M \times M \to TM$ to a function $\log\colon U \times U \subset \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$\log(q, r) = \log(\mathcal{P}(q), \mathcal{P}(r)) + r - \mathcal{P}(r) - q + \mathcal{P}(q).$$

By Lemma 1.2.9 we have

$$\log(q, r) = \mathcal{P}(r) - \mathcal{P}(q) + \mathcal{O}(d_g^2(\mathcal{P}(r), \mathcal{P}(q))) + r - \mathcal{P}(r) - q + \mathcal{P}(q) = r - q + \mathcal{O}(d_g^2(\mathcal{P}(r), \mathcal{P}(q)))$$

and therefore

$$D_q \log(q, r)|_{q=r} = -id, \tag{3.18}$$

where D_q denotes the derivative with respect to q for all $q \in M$. In the following we assume that log is smooth (i.e. C^∞) which is equivalent of M being smooth. Additionally we will only use (3.18) and the fact that Q_R is defined by (3.17). The statements would therefore also be true for any smooth function $L\colon U \subset U \times U \to \mathbb{R}^n$ satisfying $D_q l(q, r)|_{q=r} = -id$ and corresponding operator Q defined by (3.17) with log replaced by L.

Proposition 3.4.1. *Let $m \in \mathbb{N}$ and $u \in W^{m,\infty}(\Omega, M)$. Assume that*

$$\|u - Q_{\mathbb{R}^n} u\|_{L^\infty} \lesssim h^m |u|_{W^{m,\infty}} \text{ for all } u \in W^{m,\infty}(\Omega, \mathbb{R}^n). \tag{3.19}$$

Then we have

$$\|u - Q_R u\|_{L^\infty} \lesssim h^m \text{ for all } u \in W^{m,\infty}(\Omega, M), \tag{3.20}$$

where the implicit constant depends only on the norm of the derivatives of u and log.

Proof. For $q \in M$ and $U \subset \Omega$ such that $\log(q, u(x))$ is well defined for all $x \in U$ consider the function $R_q\colon U \to T_q M$ defined by $R_q(x) := \log(q, u(x))$ for all $x \in U$. By (3.19) we have

$$|Q_{\mathbb{R}^n} R_q - R_q|_{L^\infty(U)} \lesssim h^m |R_q|_{W^{m,\infty}(\Omega)}. \tag{3.21}$$

In particular $|Q_{\mathbb{R}^n} R_q|_{L^\infty}$ can be bounded by a constant independent of h and $Q_{\mathbb{R}^n} R_q$ converges to R_q for $h \to 0$. Note that by the balance law (3.17) we have $Q_{\mathbb{R}^n} R_{Q_R u(x)}(x) = 0$ for all $x \in \Omega$. Putting $q = Q_R u(x)$ in (3.21) yields that $|\log(Q_R u(x), u(x))| \lesssim h^m$ for all $x \in \Omega$ and therefore $|u(x) - Q_R u(x)| \leq |\log(Q_R u(x), u(x))| \lesssim h^m$ for all $x \in \Omega$ from which (3.20) follows. □

To prove error estimates for the derivatives we first prove that the derivatives of $Q_R u$ are bounded.

Lemma 3.4.2. *Let $m, l \in \mathbb{N}$ with $l \leq m$ and $u \in W^{m,\infty}(\Omega, M)$. Assume that for all $l' \leq l$ there exists $C_{l'} > 0$ with*

$$|u - Q_{\mathbb{R}^n} u|_{W^{l',\infty}} \leq C_{l'} h^{m-l'} |u|_{W^{l',\infty}}, \text{ for all } u \in W^{l',\infty}(\Omega, \mathbb{R}^n). \tag{3.22}$$

Then $|Q_R u|_{W^{l,\infty}}$ can be bounded by the norm of the derivatives of u and log.

Proof. Similar as in Proposition 3.4.1 we have by (3.22) for any $s \in \mathbb{N}$ and $l' \leq l$ that

$$\left| Q_{\mathbb{R}^n} D_q^s R_q - D_q^s R_q \right|_{W^{l',\infty}(U)} \lesssim h^{m-l'} |D_q^s R_q|_{W^{m,\infty}(\Omega)}, \tag{3.23}$$

where D_q^s denotes the s-th derivative by q. In particular $Q_{\mathbb{R}^n} D_q^s R_q$ can be bounded by a constant independent of h and converges to $D_q^s R_q$ for $h \to 0$. We prove the statement by induction on $l \in \mathbb{N}$. The case $l = 0$ follows from Proposition 3.4.1. By the balance law (3.17) we have $Q_{\mathbb{R}^n} R_{Q_R u(x)}(x) = 0$ and therefore also $\partial^{\vec{v}} Q_{\mathbb{R}^n} R_{Q_R u(x)}(x) = 0$ where $\vec{v} \in \mathbb{N}^s$ with $|\vec{v}|_1 = l$. On the other hand, $\partial^{\vec{v}} Q_{\mathbb{R}^n} R_{Q_R u(x)}(x)$ can be written as a sum of terms of the form

$$D_q^s \partial_y^{\vec{a}} Q_{\mathbb{R}^n} R_p(y)|_{q=Q_R u(x), y=x} \left[\partial^{\vec{b_1}} Q_R u(x), \ldots, \partial^{\vec{b_s}} Q_R u(x) \right], \tag{3.24}$$

where $\vec{a} + \sum_{i=1}^s \vec{b_i} = \vec{v}$. The term $D_q^s \partial_y^{\vec{a}} Q_{\mathbb{R}^n} R_p(y)|_{q=Q_R u(x), y=x}$ converges by (3.23) to $D_q^s \partial_y^{\vec{a}} R_p(y)|_{q=Q_R u(x), y=x}$ which can be bounded by the derivatives of u and log. For $l > 1$ and $s > 1$ the norm of (3.24) can by the induction hypothesis be bounded by the derivatives of u and log. It follows that the norm of the remaining term can also be bounded by the derivatives of u and log, i.e.

$$\left| D_q Q_{\mathbb{R}^n} R_q(x)|_{q=Q_R u(x)} \left[\partial^{\vec{v}} Q_R u(x) \right] \right| \lesssim 1.$$

By (3.23), Proposition 3.4.1 and (3.18) we have

$$\begin{aligned} \left\| D_q Q_{\mathbb{R}^n} R_q(x)|_{q=Q_R u(x)} + id \right\| &= \left\| D_q Q_{\mathbb{R}^n} R_q(x)|_{q=Q_R u(x)} - D_q R_q(x)|_{q=Q_R u(x)} \right\| \\ &\quad + \left\| D_q R_q(x)|_{q=Q_R u(x)} - D_q R_q(x)|_{q=u(x)} \right\| \\ &\quad + \left\| D_q R_q(x)|_{q=u(x)} + id \right\| \\ &\lesssim h^m. \end{aligned}$$

Hence,

$$\begin{aligned}|\partial^{\vec{v}} Q_R u(x)| &\leq \left|(D_q Q_{\mathbb{R}^n} R_q(x)|_{q=Q_R u(x)} + id)\left[\partial^{\vec{v}} Q_R u(x)\right]\right| \\ &\quad + \left|D_q Q_{\mathbb{R}^n} R_q(x)|_{q=Q_R u(x)}\left[\partial^{\vec{v}} Q_R u(x)\right]\right| \\ &\lesssim h^m |\partial^{\vec{v}} Q_R u(x)| + 1,\end{aligned}$$

which yields the desired result. □

We can now prove an error estimate for the derivatives. We assume that $|\phi_i|_{W^{l,\infty}} \lesssim h^{-l}$, which is true if the basis functions scale quasi-uniformly with the mesh width h.

Theorem 3.4.3. *Consider the same setting as in Lemma 3.4.2. Assume further that $|\phi_i|_{W^{l,\infty}} \lesssim h^{-l}$. Then we have*

$$|u - Q_R u|_{W^{l,\infty}} \lesssim h^{m-l},$$

where the implicit constant depends only on the derivatives of u and log.

Proof. The case $l = 0$ is Proposition 3.4.1. For $l \geq 1$ consider the function $v = u - Q_R u$. Its derivatives can by Lemma 3.4.2 be bounded by the derivatives of u and log. Hence, we have by (3.22), the assumption $|\phi_i|_{W^{l,\infty}} \lesssim h^{-l}$ and Proposition 3.4.1 that

$$\begin{aligned}|u - Q_R u|_{W^{l,\infty}} &= |v|_{W^{l,\infty}} \\ &\leq |v - Q_{\mathbb{R}^n} v|_{W^{l,\infty}} + |Q_{\mathbb{R}^n} v|_{W^{l,\infty}} \\ &\lesssim h^{m-l} + \left|\sum_{i\in I} \phi_i (u(\xi_i) - Q_R u(\xi_i))\right|_{W^{l,\infty}} \\ &\lesssim h^{m-l} + \max_{i\in I} |\phi_i|_{W^{l,\infty}} |u(\xi_i) - Q_R u(\xi_i)| \\ &\lesssim h^{m-l}. \quad \square\end{aligned}$$

So far, we did not specify the dependence of our estimate of $\|u - Q_R u\|_{W^{l,p}}$ in terms of the derivatives of u. It is for example still unclear if Theorem 3.3.4 with $Q_{\mathcal{P}}$ replaced by Q_R holds. If log is replaced by L defined by $L(p, q) = P_{T_pM}(q - p)$ one can prove the corresponding statement. In the next proposition, we show that for some specific cases we have the same error bound as in the linear theory. To get this result, we compare the projection average based approximation with the Riemannian average based approximation.

Proposition 3.4.4. *Consider the same setting as in Theorem 3.3.4 and assume further that $m \leq 2$ and $u \in W^{1,\infty}$. Then for h sufficiently small we have*

$$\|u - Q_R u\|_{L^p} \lesssim h^m |u|_{W^{m,p}}.$$

Proof. Using the triangle inequality, Theorem 3.3.4 and Proposition 1.4.5 we have

$$\|u - Q_R u\|_{L^p} \leq \|u - Q_{\mathcal{P}} u\|_{L^p} + \|Q_{\mathcal{P}} u - Q_R u\|_{L^p} \lesssim h^m |u|_{W^{m,p}} + h^3 |u|^3_{W^{1,\infty}}.$$

As $m \leq 2 < 3$ we can bound the first term by the second term for sufficiently small h. □

3.5 Error estimates for approximation operators with B-splines

In the previous sections, the underlying assumption was that the approximation operator $Q_{\mathbb{R}} f = \sum_{i \in I} \phi f(\xi)$ is exact for polynomials up to a certain degree. This property is crucial to get the corresponding error bounds. In this section, we consider a case where $Q_{\mathbb{R}}$ is only exact for polynomials of degree 1 but where it is possible to reproduce polynomials of higher degree using linear combinations. Our task will be to generalize a corresponding approximation operator to the case of manifold-valued functions.

We consider the case where $\Omega = \mathbb{R}$ and $(\phi_i)_{i \in I}$ are B-splines. In Section 3.5.1, we introduce uniform B-splines and present approximation operators with optimal approximation order for the linear case (i.e. when $M = \mathbb{R}^n$). A natural generalization of such an approximation operator is presented in Section 3.5.2. In Section 3.5.3, we present an approximation operator into the projection average based function space which has the same approximation order as its linear analog. Finally, in Section 3.5.4, we introduce B-splines for arbitrary knots and show how to compute the L^2 best approximation onto $V_{\mathcal{P}}$. We will see that for nonuniform B-splines we can in general not have the same convergence order as in the linear case.

3.5.1 Linear theory

We first define uniform B-splines.

Definition 3.5.1. The uniform B-splines are recursively defined by

$$B_0(x) := \begin{cases} 1 & -0.5 \leq x < 0.5 \\ 0 & \text{otherwise.} \end{cases} \quad \text{and} \quad B_k(x) := \int_{x-0.5}^{x+0.5} B_{k-1}(y) dy \ \text{ for all } x \in \mathbb{R}, k \in \mathbb{N}_{>0}.$$

The following basic properties of uniform B-splines are important for us

i) $B_k \geq 0$.

ii) $\operatorname{supp} B_k = \left[-\frac{k+1}{2}, \frac{k+1}{2}\right]$.

iii) $\int_{\mathbb{R}} B_k(x) dx = 1$.

iv) $\sum_{i\in\mathbb{Z}} B_k(i) = 1$.

v) for all $i \in \{0,\dots,k\}$ we have that the restriction of B_k to the interval $\left[-\frac{k+1}{2}+i, -\frac{k+1}{2}+i+1\right]$ is a nontrivial polynomial of degree k.

vi) $B_k \in C^{k-1}$.

To compute uniform B-splines, one can use the three point recursion formula

$$B_k(x) = \frac{\frac{k+1}{2}+x}{k} B_{k-1}(x+0.5) + \frac{\frac{k+1}{2}-x}{k} B_{k-1}(x-0.5).$$

For $h > 0$ and $k \in \mathbb{N}$ the set of basis functions we use is

$$\Phi := (\phi_i)_{i\in\mathbb{Z}} \quad \text{where} \quad \phi_i(x) := B_k(h^{-1}x - i), \text{ for all } x \in \mathbb{R}, \text{ and } i \in \mathbb{Z}.$$

Unfortunately, the operator $Q_{\mathbb{R}^n}\colon C(\mathbb{R},\mathbb{R}^n) \to V_{\mathbb{R}^n}$, i.e. $Q_{\mathbb{R}^n}u = \sum_{i\in\mathbb{Z}} \phi_i u(hi)$ is only exact for polynomials of degree 1 which is not optimal.

An interpolation operator

In this section, we construct an interpolation operator $I_{\mathbb{R}^n}\colon C(\mathbb{R},\mathbb{R}^n) \to V_{\mathbb{R}^n}$, i.e. an operator which satisfies $I_{\mathbb{R}^n}u(ih) = u(hi)$ for all $u \in C(\mathbb{R},\mathbb{R}^n)$ and $i \in \mathbb{Z}$. Before we start we need to introduce some notation. We consider the sampling operator $S\colon C(\mathbb{R},\mathbb{R}^n) \to (\mathbb{R}^n)^{\mathbb{Z}}$ defined by

$$(Su)_i := u(hi) \quad \text{for all } u \in C(\mathbb{R},\mathbb{R}^n) \text{ and } i \in \mathbb{Z}.$$

We indicate sequences with fraktur letters. For sequences $\mathfrak{a} \in \mathbb{R}^{\mathbb{Z}}$ and $\mathfrak{b} \in (\mathbb{R}^n)^{\mathbb{Z}}$ we define the convolution $\mathfrak{a} * \mathfrak{b} \in (\mathbb{R}^n)^{\mathbb{Z}}$ by

$$(\mathfrak{a} * \mathfrak{b})_i := \sum_{j\in\mathbb{Z}} \mathfrak{a}_j \mathfrak{b}_{i-j} \quad \text{for all } i \in \mathbb{Z}.$$

To construct an interpolation operator we choose the nodes $\mathfrak{p} = (p_i)_{i\in\mathbb{Z}} \subset \mathbb{R}^n$ in such a way that $\mathfrak{b} * \mathfrak{p} = Su$, where $\mathfrak{b} \in \mathbb{R}^{\mathbb{Z}}$ is defined by $\mathfrak{b}_i := B_k(i)$ for all $i \in \mathbb{Z}$. In Lemma 3.5.2 we show that $\mathfrak{b}$ has an inverse $\mathfrak{b}^{-1} \in \mathbb{R}^{\mathbb{Z}}$ with respect to the convolution "*". Then we can simply set $\mathfrak{p} := \mathfrak{b}^{-1} * Su$ and define

$$I_{\mathbb{R}^n} := Q_{\mathbb{R}^n} \circ \mathfrak{b}^{-1} \circ S,$$

where by abuse of notation $\mathfrak{b}^{-1}$ denotes the operator $x \mapsto \mathfrak{b}^{-1} * x$ and $Q_{\mathbb{R}^n}$ is the operator $\mathfrak{p} \mapsto \sum_{i\in\mathbb{Z}} \mathfrak{p}_i \phi_i$. Then we have

$$SI_{\mathbb{R}^n}u = \mathfrak{b} * \mathfrak{b}^{-1} * Su = Su.$$

Hence, $I_{\mathbb{R}^n}$ is an interpolation operator. Let us now prove that $\mathfrak{b}$ has an inverse.

Lemma 3.5.2. *The sequence $\mathfrak{b}$ has an inverse, i.e. there exists a sequence $\mathfrak{b}^{-1} \in \mathbb{R}^{\mathbb{Z}}$ with*

$$(\mathfrak{b}^{-1} * \mathfrak{b})_i = \begin{cases} 1 & i = 0 \\ 0 & i \neq 0. \end{cases} \tag{3.25}$$

Proof. We consider the Eulerian polynomial divided by the monomial $z^{\frac{k+1}{2}}$, i.e.

$$A(z) := k! \sum_{j \in \mathbb{Z}} \mathfrak{b}_j z^j.$$

If A is positive on the unit circle, i.e. if $A(z) > 0$ for all $z \in \mathbf{C}$ with $|z| = 1$, there exists by the lemma of Wiener [55] a sequence $\mathfrak{b}^{-1} \in \mathbb{R}^{\mathbb{Z}}$ such that $\|\mathfrak{b}^{-1}\|_{\ell^1} < \infty$ and

$$\sum_{j \in \mathbb{Z}} \mathfrak{b}_j^{-1} z^j = \frac{1}{A(z)} \text{ for all } z \in \mathbf{C} \text{ with } |z| = 1. \tag{3.26}$$

Then we have

$$1 = A(z)\frac{1}{A(z)} = \left(\sum_{j=-S}^{S} \mathfrak{b}_j z^j \right) \left(\sum_{j \in \mathbb{Z}} \mathfrak{b}_j^{-1} z^j \right) = \sum_{i \in \mathbb{Z}} \left(\sum_{j=-S}^{S} \mathfrak{b}_j \mathfrak{b}_{i-j}^{-1} \right) z^i = \sum_{i \in \mathbb{Z}} (\mathfrak{b} * \mathfrak{b}^{-1})_i z^i$$

from which (3.25) follows. It is left to prove that A is positive on the unit circle. As $\mathfrak{b}_i = \mathfrak{b}_{-i}$ for all $i \in \mathbb{Z}$ we have that A is real-valued on the unit circle. Furthermore by continuity of A and $A(1) = 1$ it suffices to prove that $A(z) \neq 0$ for all $z \in \mathbf{C}$ with $|z| = 1$. By [22] all roots of A are real-valued, hence it suffices to prove that $A(-1) \neq 0$. From [13] we have for $k > 0$ that $A_k(-1) = (4^k - 2^k)\zeta(-(k-1)) \neq 0$ where ζ is the Riemann–Zeta function. □

The sequence $\mathfrak{b}^{-1}$ has, as opposed to the sequence $\mathfrak{b}$, infinitely many nonzero entries. As a result, $I_{\mathbb{R}^n}$ is not a local operator, i.e. $I_{\mathbb{R}^n} u(x)$ depends not only on u in a neighborhood of $x \in \mathbb{R}$. However, we can show that the entries of $\mathfrak{b}^{-1}$ decay exponentially. The consequence is that the dependence of $I_{\mathbb{R}^n} u(x)$ on $u(y)$ decays exponentially with the distance $|x - y|$.

Lemma 3.5.3. *There exists $c \in (0, 1)$ such that we have*

$$|\mathfrak{b}_j^{-1}| \lesssim c^{|j|} \text{ for all } j \in \mathbb{Z}.$$

Proof. By continuity of A there exists a constant $c \in (0, 1)$ such that A does not have any zeros in the annulus $\{z \in \mathbf{C} \mid c < |z| < c^{-1}\}$. By standard complex analysis we have

$$\mathfrak{b}_j^{-1} = \int_{B_0(1)} \frac{z^{-j}}{A(z)} dz = \int_{B_0(r)} \frac{z^{-j}}{A(z)} dz$$

for any $r \in [c, c^{-1}]$ where $B_0(r) = \{z \in \mathbf{C} \mid |z| = r\}$. Hence

$$\begin{array}{llll} |\mathfrak{b}_j^{-1}| & = \int_{B_0(c^{-1})} \frac{z^{-j}}{A(z)} dz & \leq c^j \int_{B_0(c^{-1})} \frac{1}{|A(z)|} dz & \lesssim c^{|j|} \quad \text{for } j \geq 0 \text{ and} \\ |\mathfrak{b}_j^{-1}| & = \int_{B_0(c)} \frac{z^{-j}}{A(z)} dz & \leq c^{-j} \int_{B_0(c)} \frac{1}{|A(z)|} dz & \lesssim c^{|j|} \quad \text{for } j < 0. \quad \square \end{array}$$

We have the following error estimate for $I_{\mathbb{R}^n}$.

Theorem 3.5.4. *Let $m \in \mathbb{N}_{>0}$ and $l \leq \min(k, m)$. Then we have*

$$|u - I_{\mathbb{R}^n} u|_{W^{l,\infty}} \lesssim h^{\min(m,k+1)-l} |u|_{W^{\min(m,k+1),\infty}} \tag{3.27}$$

for all $u \in W^{m,\infty}(\mathbb{R}, \mathbb{R})$ and $h > 0$.

Proof. We first prove exactness of $I_{\mathbb{R}}$ for polynomials of degree smaller or equal to k. For $\mathfrak{u} \in (\mathbb{R}^n)^{\mathbb{Z}}$ we define the discrete derivative $\nabla\mathfrak{u} \in (\mathbb{R}^n)^{\mathbb{Z}}$ by $(\nabla\mathfrak{u})_i := \mathfrak{u}_{i+1} - \mathfrak{u}_i$ for all $i \in \mathbb{Z}$. By the definition of uniform B-splines (Definition 3.5.1) we have

$$B_k'(x) = B_{k-1}\left(x + \frac{1}{2}\right) - B_{k-1}\left(x - \frac{1}{2}\right).$$

Hence,

$$\begin{aligned}
\frac{s}{dx} I_{\mathbb{R}} u(x) &= \sum_{i\in\mathbb{Z}} (\mathfrak{b}^{-1} S u)_i B_k'(x - i) \\
&= \sum_{i\in\mathbb{Z}} (\mathfrak{b}^{-1} S u)_i \left(B_{k-1}\left(x - i + \frac{1}{2}\right) - B_{k-1}\left(x - i - \frac{1}{2}\right)\right) \\
&= \sum_{i\in\mathbb{Z}} ((\mathfrak{b}^{-1} S u)_{i+1} - (\mathfrak{b}^{-1} S u)_i) B_{k-1}\left(x - \frac{1}{2} - i\right) \\
&= \sum_{i\in\mathbb{Z}} (\mathfrak{b}^{-1} \nabla S u)_i B_{k-1}\left(x - \frac{1}{2} - i\right).
\end{aligned}$$

Inductively it follows that

$$\frac{d^l}{dx^l} I_{\mathbb{R}} u(x) = \sum_{i\in\mathbb{Z}} (\mathfrak{b}^{-1} \nabla^l S u)_i B_{k-l}\left(x - \frac{l}{2} - i\right).$$

Let now p be a polynomial of degree smaller or equal to $k \in \mathbb{N}$. Note that $\nabla^k S p$ is then constant. Hence it follows that $\frac{d^k}{dx^k} I_{\mathbb{R}} p$ is constant. Therefore, $I_{\mathbb{R}} p$ is a polynomial of degree at most k and since p and $I_{\mathbb{R}} p$ coincide on more than k points we have $I_{\mathbb{R}} p = p$. Hence, $I_{\mathbb{R}}$ is exact for polynomials of degree smaller or equal to k.

Consider now $x \in \mathbb{R}$ and let p be the Taylor expansion of u at x of degree $m - 1$, i.e.

$$p(y) := \sum_{j=0}^{m-1} \frac{d^j}{dx^j} u(x) \frac{(y - x)^j}{j!}.$$

Note that we have

$$\left|\frac{d^l}{dx^l}(u - p)(y)\right| \lesssim |u|_{W^{\min(m,k+1),\infty}} |y - x|^{\min(m,k+1)-l},$$

and hence

$$\left|(\nabla^l(S(u-p)))_i\right| \lesssim h^{\min(m,k+1)-l}|i-h^{-1}x|^{\min(m,k+1)-l}|u|_{W^{\min(m,k+1),\infty}}.$$

We now have

$$\begin{aligned}
\left|\frac{d^l}{dx^l}(u-I_{\mathbb{R}}u)(x)\right| &= \left|\frac{d^l}{dx^l}(u-p-I_{\mathbb{R}}(u-p)(x)\right| \\
&= \left|\frac{d^l}{dx^l}I_{\mathbb{R}}(u-p)(x)\right| \\
&= \left|\sum_{i\in\mathbb{Z}}(\mathfrak{b}^{-1}\nabla^l S(u-p))_i B_{k-l}\left(x-\frac{l}{2}-i\right)\right| \\
&\lesssim \sum_{i\in\mathbb{Z}} h^{\min(m,k+1)-l}B_{k-l}\left(x-\frac{l}{2}-i\right)|u|_{W^{\min(m,k+1),\infty}} \\
&\lesssim h^{\min(m,k+1)-l}|u|_{W^{\min(m,k+1),\infty}}. \quad \square
\end{aligned}$$

Quasi-interpolation operators

We consider also local approximation operators of the form

$$QI_{\mathbb{R}^n} := Q_{\mathbb{R}^n}\circ\mathfrak{c}\circ S$$

where $\mathfrak{c}\in\mathbb{R}^{\mathbb{Z}}$ has only finitely many nonzero entries. It can be shown that for all $k\in\mathbb{N}$ there exists $\mathfrak{c}=(c_i)_{i\in\mathbb{Z}}\in\mathbb{R}^{\mathbb{Z}}$ with finite support such that $QI_{\mathbb{R}^n}$ is exact for polynomials of degree less or equal to k. For example for $k=3$ and $k=5$ one can choose

$$(\mathfrak{c}_{-1},\mathfrak{c}_0,\mathfrak{c}_1)=\frac{1}{6}(-1,8,-1)\quad\text{and}\quad(\mathfrak{c}_{-2},\ldots,\mathfrak{c}_2)=\frac{1}{240}(13,-112,438,-112,13).$$

Due to the polynomial exactness of $QI_{\mathbb{R}^n}$ the following approximation error estimate can be shown.

Theorem 3.5.5. *Assume that $\mathfrak{c}\in\mathbb{R}^{\mathbb{Z}}$ has finite support, $QI_{\mathbb{R}}$ is exact for polynomials of degree less or equal to $r\in\mathbb{N}$, $m\in\mathbb{N}_{>0}$ and $l\le\min(k,m,r+1)$. Then we have*

$$|u-QI_{\mathbb{R}^n}u|_{W^{l,\infty}} \lesssim h^{\min(m,r+1)-l}|u|_{W^{\min(m,r+1),\infty}} \tag{3.28}$$

for all $u\in W^{m,\infty}(\mathbb{R}^n)$ and $h>0$.

3.5.2 The naive generalization of the quasi-interpolation operator

Let now M be a Riemannian submanifold of $\mathbb{R}^n$ and $u\colon\mathbb{R}\to M\subset\mathbb{R}^n$. Since in general $(\mathfrak{c}*Su)_i$ is not in M we have in general that $Q_{\mathcal{P}}\mathfrak{c}Su$ is not in $V_{\mathcal{P}}$. In this section, we introduce a natural approximation operator into $V_{\mathcal{P}}$ by replacing $\mathfrak{c}*Su$ with $\mathcal{P}(\mathfrak{c}*Su)$.

Definition 3.5.6. Let $QI_{\mathcal{P}} \colon C(\mathbb{R}, M) \to V$ be defined by

$$QI_{\mathcal{P}} := Q_{\mathcal{P}} \mathcal{P} \mathfrak{c} S.$$

Unfortunately, this operator does not have the optimal convergence order for $k > 4$ as can be seen in numerical experiments. In this section, we show that the operator has a convergence order of at least $\min(m, 2) - l$. Numerical experiments suggest that the actual convergence order is $\min(m, 4) - l$.

The exactness of $QI_{\mathcal{P}}$ also depend on the polynomial exactness of $\mathfrak{c}$.

Definition 3.5.7. We say that $\mathfrak{c} \in \mathbb{R}^{\mathbb{Z}}$ is exact for polynomials of degree less or equal to $r \in \mathbb{N}$ if $\mathfrak{c} * Sq = Sq$ for all polynomials q of degree less or equal to r.

The next theorem shows that we can not have simultaneously polynomial exactness of $\mathfrak{c}$ and $QI_{\mathbb{R}^n}$ of high order.

Theorem 3.5.8. *For $h > 0$ there is no sequence $\mathfrak{c} \in \mathbb{R}^{\mathbb{Z}}$ and $k \in \mathbb{N}$ such that both $\mathfrak{c} \in \mathbb{R}^{\mathbb{Z}}$ and $QI_{\mathbb{R}^n}$ are exact for polynomials of degree less or equal to 2.*

Proof. Suppose that $\mathfrak{c} \in \mathbb{R}^{\mathbb{Z}}$ and $QI_{\mathbb{R}^n}$ are exact for polynomials of degree less or equal to 2. Let $a \in \mathbb{R} \backslash \{hi | i \in \mathbb{Z}\}$ and $q_a(x) = (x - a)^2$ for all $x \in \mathbb{R}$. Note that we can assume $k > 0$ and therefore also $B_k(x) > 0$ for all $x \in (0, 1)$. By polynomial exactness of $QI_{\mathbb{R}^n}$ and the non negativity of B-splines we have

$$\begin{aligned} 0 &= q(a) = QI_{\mathbb{R}^n} q(a) \\ &= \sum_{i \in \mathbb{Z}} (\mathfrak{c} * Sq)_i B_k(h^{-1}a - i) \\ &= \sum_{i \in \mathbb{Z}} (Sq)_i B_k(h^{-1}a - i) \\ &\geq (Sq)_{\lceil h^{-1}a \rceil} B_k(h^{-1}a - \lceil h^{-1}a \rceil) \\ &> 0, \end{aligned}$$

which is a contradiction. □

For $q \in [1, \infty]$ we define the ℓ^q-norm of $\mathfrak{u} = (\mathfrak{u}_i)_{i \in \mathbb{Z}} \in (\mathbb{R}^n)^{\mathbb{Z}}$ by

$$\|\mathfrak{u}\|_{\ell^q} := \begin{cases} (\sum_{i \in \mathbb{Z}} |\mathfrak{u}_i|^q)^{\frac{1}{q}} & \text{if } q \in [1, \infty) \\ \sup_{i \in \mathbb{Z}} |\mathfrak{u}_i| & \text{if } q = \infty. \end{cases}$$

Lemma 3.5.9. *Assume that $m > 0$ and that $\mathfrak{c} \in \mathbb{R}^{\mathbb{Z}}$ has finite support, i.e. $\mathfrak{c}_i \neq 0$ only for finitely many $i \in \mathbb{Z}$, and is exact for polynomials of degree 1. Then we have*

$$\|\mathcal{P}(\mathfrak{c} * Su) - \mathfrak{c} * Su\|_{\ell^\infty} \lesssim h^{\min(m,2)} |u|_{W^{\min(m,2),\infty}},$$

for all $u \in W^{m,\infty}(\mathbb{R}, M)$ and h small enough.

Proof. Let $K > 0$ be such that the support of $\mathfrak{c}$ is in $\{-K, \dots, K\}$. By polynomial exactness of $\mathfrak{c}$ of order 1, and the Bramble-Hilbert lemma, we have for any $i \in \mathbb{Z}$

$$\begin{aligned} |(\mathfrak{c} * Su)_i - (Su)_i| &= \inf_{q \in \mathcal{P}_1} |(\mathfrak{c} * Su)_i - (\mathfrak{c} * Sq)_i + (Sq)_i - (Su)_i| \\ &\leq \inf_{q \in \mathcal{P}_1} |(\mathfrak{c} * S(u-q))_i| + |(S(u-q))_i| \\ &\leq \|\mathfrak{c}\|_{\ell^1} \inf_{q \in \mathcal{P}_1} \|u - q\|_{L^\infty([h(i-K)\ h(i+K)])} \\ &\lesssim h^{\min(m,2)} |u|_{W^{\min(m,2),\infty}} \end{aligned}$$

By the triangle inequality, $\mathcal{P}(Su) = Su$ and $|\mathcal{P}(x) - \mathcal{P}(y)| \leq \mathrm{Lip}(\mathcal{P})|x - y|$ where $\mathrm{Lip}(\mathcal{P})$ denotes the Lipschitz constant of $\mathcal{P}$ we have

$$\begin{aligned} |\mathcal{P}((\mathfrak{c} * Su)_i) - (\mathfrak{c} * Su)_i| &= |\mathcal{P}((\mathfrak{c} * Su)_i) - \mathcal{P}((Su)_i) + (Su)_i - (\mathfrak{c} * Su)_i| \\ &\leq |\mathcal{P}((\mathfrak{c} * Su)_i) - \mathcal{P}((Su)_i)| + |(Su)_i - (\mathfrak{c} * Su)_i| \\ &\lesssim |(\mathfrak{c} * Su)_i - (Su)_i| \\ &\leq h^{\min(m,2)} |f|_{W^{\min(m,2),\infty}}. \quad \square \end{aligned}$$

We can now prove an approximation result for $QI_{\mathcal{P}}$.

Theorem 3.5.10. *Let $l \leq \min(m, 2)$ and assume that $c \in \mathbb{R}^{\mathbb{Z}}$ has finite support and is exact for polynomials of degree 1. Furthermore, assume that $Q_{\mathbb{R}}$ is also exact for polynomials of degree less or equal to 1. Then for $u \in W^{m,\infty}(\mathbb{R}, M)$ and h small enough we have*

$$|u - QI_{\mathcal{P}} u|_{W^{l,\infty}} \lesssim h^{\min(m,2)-l} |u|_{W^{\min(m,2),\infty}}.$$

Proof. Using the definition of $QI_{\mathcal{P}}$ we have

$$|u - QI_{\mathcal{P}} u|_{W^{l,\infty}} = \left| u - \mathcal{P}\left(\sum_{i \in \mathbb{Z}} \mathcal{P}((\mathfrak{c} * Su)_i) \phi_i \right) \right|_{W^{l,\infty}}.$$

By Lipschitz continuity of $\mathcal{P}$ by Section 3.2 it is enough to prove that

$$\left| u - \sum_{i \in \mathbb{Z}} \mathcal{P}((\mathfrak{c} * Su)_i) \phi_i \right|_{W^{l,\infty}} \lesssim h^{\min(m,2)-l}.$$

Note that for $x \in \mathbb{R}$ the number of $i \in \mathbb{Z}$ such that x is in the support of ϕ_i is bounded by a constant depending only on k. Furthermore we have for any $i \in \mathbb{Z}$ that $|\phi_i|_{W^{l,\infty}} = |\phi_0|_{W^{l,\infty}} \lesssim h^{-l}$. Using Lemma 3.5.9 we get

$$\begin{aligned} \left| u - \sum_{i \in \mathbb{Z}} \mathcal{P}((\mathfrak{c} * Su)_i) \phi_i \right|_{W^{l,\infty}} &\lesssim \left| u - Q_{\mathbb{R}^n} u + \sum_{i \in \mathbb{Z}} ((\mathfrak{c} * Su)_i - \mathcal{P}((\mathfrak{c} * Su)_i)) \phi_i \right|_{W^{l,\infty}} \\ &\lesssim |u - Q_{\mathbb{R}^n} u|_{W^{l,\infty}} + \|\mathfrak{c} * Su - \mathcal{P}(\mathfrak{c} * Su)\|_{\ell^\infty} |\phi_0|_{W^{l,\infty}} \\ &\lesssim h^{\min(m,2)-l} |u|_{W^{\min(m,2),\infty}} + h^{\min(m,2)} h^{-l} |u|_{W^{\min(m,2),\infty}} \\ &\lesssim h^{\min(m,2)-l} |u|_{W^{\min(m,2),\infty}}. \quad \square \end{aligned}$$

3.5.3 Generalization of the interpolation operator

As we have seen in Section 3.5.2, the naive approximation operator is not optimal and the question remains if there exists an optimal approximation operator into $V_{\mathcal{P}}$ respectively V_R. In this section we consider the generalization $I_{\mathcal{P}} \colon C(\mathbb{R}, M) \to V_{\mathcal{P}}$ defined by

$$I_{\mathcal{P}} := Q_{\mathcal{P}} \circ (S \circ Q_{\mathcal{P}})^{-1} \circ S = \mathcal{P} \circ Q_{\mathbb{R}^n} \circ (\mathcal{P} \circ \mathfrak{b})^{-1} \circ S, \tag{3.29}$$

where $(\mathcal{P} \circ \mathfrak{b})^{-1} \colon M^{\mathbb{Z}} \to M^{\mathbb{Z}}$ is the inverse (which as we will show in Lemma 3.5.12 exists) of the operator $\mathcal{P}\mathfrak{b} := \mathcal{P} \circ \mathfrak{b} \colon M^{\mathbb{Z}} \to M^{\mathbb{Z}}$. We show that the interpolation operator $I_{\mathcal{P}}$ has the optimal approximation order.

Theorem 3.5.11. *Let $m \in \mathbb{N}_{>0}$, $l \leq \min(k, m)$, $u \in W^{m,\infty}$ and $C > 0$ larger than the implicit constant of Theorem 3.5.4. Then for h small enough we have*

$$|u - I_{\mathcal{P}} u|_{W^{l,\infty}} \leq C h^{\min(m,k+1)-l} |u|_{W^{\min(m,k+1),\infty}}. \tag{3.30}$$

Let us first prove the existence of $(\mathcal{P}\mathfrak{b})^{-1}$. The idea is to define for $\mathfrak{u} \in M^{\mathbb{Z}}$ a retraction $R \colon M^{\mathbb{Z}} \to M^{\mathbb{Z}}$ where the solution $\mathfrak{v}$ of $\mathcal{P}\mathfrak{b}\mathfrak{v} = \mathfrak{u}$ is the fixpoint of R. Then we can use Banach's fixpoint theorem to prove existence and uniqueness.

Lemma 3.5.12. *Let $u \colon \mathbb{R} \to M$ be a Lipschitz continuous function and $\mathfrak{u} := Su$. For h small enough there exists a unique $\mathfrak{v} \in M^{\mathbb{Z}}$ such that*

$$\mathcal{P}\mathfrak{b}\mathfrak{v} = \mathfrak{u}, \quad \text{and} \quad \|\mathfrak{v} - \mathfrak{u}\|_{\ell^\infty} \lesssim h. \tag{3.31}$$

Proof. We define a retraction $R \colon M^{\mathbb{Z}} \to M^{\mathbb{Z}}$ by

$$R(\mathfrak{v}) := \mathcal{P}\left(\mathfrak{v} - \mathfrak{b}^{-1}\mathcal{P}\mathfrak{b}\mathfrak{v} + \mathfrak{b}^{-1}\mathfrak{u}\right).$$

Note that every solution of (3.31) is a fixpoint of R. For $\mathfrak{v}^0, \mathfrak{v}^1 \in M^{\mathbb{Z}}$ with $\|\mathfrak{u} - \mathfrak{v}^0\|_{\ell^\infty} \lesssim h$ and $\|\mathfrak{u} - \mathfrak{v}^1\|_{\ell^\infty} \lesssim h$ we have

$$\begin{aligned}
\|R(\mathfrak{v}^1) - R(\mathfrak{v}^0)\|_{\ell^\infty} &= \left\|\mathcal{P}\left(\mathfrak{v}^1 - \mathfrak{b}^{-1}\mathcal{P}\mathfrak{b}\mathfrak{v}^1 + \mathfrak{b}^{-1}\mathfrak{u}\right) - \mathcal{P}\left(\mathfrak{v}^0 - \mathfrak{b}^{-1}\mathcal{P}\mathfrak{b}\mathfrak{v}^0 + \mathfrak{b}^{-1}\mathfrak{u}\right)\right\|_{\ell^\infty} \\
&\leq \operatorname{Lip}(\mathcal{P}) \left\|\left(\mathfrak{v}^1 - \mathfrak{b}^{-1}\mathcal{P}\mathfrak{b}\mathfrak{v}^1 + \mathfrak{b}^{-1}\mathfrak{u}\right) - \left(\mathfrak{v}^0 - \mathfrak{b}^{-1}\mathcal{P}\mathfrak{b}\mathfrak{v}^0 + \mathfrak{b}^{-1}\mathfrak{u}\right)\right\|_{\ell^\infty} \\
&\leq \operatorname{Lip}(\mathcal{P}) \left\|\mathfrak{b}^{-1}\right\|_{\ell^\infty} \left\|\mathfrak{b}\mathfrak{v}^1 - \mathfrak{b}\mathfrak{v}^0 - \left(\mathcal{P}\mathfrak{b}\mathfrak{v}^1 - \mathcal{P}\mathfrak{b}\mathfrak{v}^0\right)\right\|_{\ell^\infty}.
\end{aligned}$$

For $t \in [0, 1]$ define $\mathfrak{v}(t) := (1-t)\mathfrak{v}^0 + t\mathfrak{v}^1$. Then we have

$$\mathcal{P}\mathfrak{b}\mathfrak{v}^1 - \mathcal{P}\mathfrak{b}\mathfrak{v}^0 = \int_0^1 \mathcal{P}'(\mathfrak{b}\mathfrak{v}(t))\mathfrak{b}(\mathfrak{v}^1 - \mathfrak{v}^0)dt = \int_0^1 \mathcal{P}'(\mathfrak{b}\mathfrak{v}(t))dt\mathfrak{b}(\mathfrak{v}^1 - \mathfrak{v}^0).$$

By the smoothness of $\mathcal{P}$ and Lemma B.2 there exists $\mathfrak{a} \in \mathbb{R}^{\mathbb{Z}}$ with

$$\left\|\int_0^1 \mathcal{P}'(\mathfrak{b}\mathfrak{v}(t))dt - \frac{\mathcal{P}'(\mathfrak{v}^0) + \mathcal{P}'(\mathfrak{v}^1)}{2}\right\|_{\ell^\infty} \lesssim \left\|\int_0^1 \mathfrak{b}\mathfrak{v}(t)dt - \frac{\mathfrak{v}^0 + \mathfrak{v}^1}{2}\right\|_{\ell^\infty} + \left\|\mathfrak{v}^0 - \mathfrak{v}^1\right\|_{\ell^\infty}^2$$

$$\begin{aligned} &= \frac{1}{2}\left\|\mathfrak{b}\mathfrak{a}\nabla(\mathfrak{v}^0+\mathfrak{v}^1)\right\|_{\ell^\infty} + \left\|\mathfrak{v}^0-\mathfrak{v}^1\right\|_{\ell^\infty}^2 \\ &\lesssim h + \left\|\mathfrak{v}^0-\mathfrak{v}^1\right\|_{\ell^\infty}^2. \end{aligned}$$

Lemma A.1 yields

$$\left\|\mathfrak{v}^1-\mathfrak{v}^0-\left(\frac{\mathcal{P}'(\mathfrak{v}^0)+\mathcal{P}'(\mathfrak{v}^1)}{2}\right)(\mathfrak{v}^1-\mathfrak{v}^0)\right\|_{\ell^\infty} \lesssim \|\mathfrak{v}^1-\mathfrak{v}^0\|_{\ell^\infty}^2$$

By Lemma B.2 we have

$$\|(\mathfrak{v}^1-\mathfrak{v}^0)-\mathfrak{b}(\mathfrak{v}^1-\mathfrak{v}^0)\|_{\ell^\infty} \lesssim \left\|\nabla\left(\mathfrak{v}^1-\mathfrak{v}^0\right)\right\|_{\ell^\infty} \lesssim \|\mathfrak{v}^1-\mathfrak{v}^0\|_{\ell^\infty}^2$$

The triangle inequality and the previous estimates yield

$$\begin{aligned} \|R(\mathfrak{v}^1)-R(\mathfrak{v}^0)\|_{\ell^\infty} &\lesssim \left\|\mathfrak{b}\mathfrak{v}^1-\mathfrak{b}\mathfrak{v}^0-\left(\mathcal{P}\mathfrak{b}\mathfrak{v}^1-\mathcal{P}\mathfrak{b}\mathfrak{v}^0\right)\right\|_{\ell^\infty} \\ &= \left\|\mathfrak{b}\mathfrak{v}^1-\mathfrak{b}\mathfrak{v}^0-\int_0^1\mathcal{P}'(\mathfrak{b}\mathfrak{v}(t))dt\mathfrak{b}(\mathfrak{b}\mathfrak{v}^1-\mathfrak{b}\mathfrak{v}^0)\right\|_{\ell^\infty} \\ &\lesssim \left\|(\mathfrak{b}\mathfrak{v}^1-\mathfrak{b}\mathfrak{v}^0)-(\mathfrak{v}^1-\mathfrak{v}^0)\right\|_{\ell^\infty} \\ &\quad+\left\|(\mathfrak{v}^1-\mathfrak{v}^0)-\left(\frac{\mathcal{P}'(\mathfrak{v}^0)+\mathcal{P}'(\mathfrak{v}^1)}{2}\right)(\mathfrak{v}^1-\mathfrak{v}^0)\right\|_{\ell^\infty} \\ &\quad+\left\|\left(\frac{\mathcal{P}'(\mathfrak{v}^0)+\mathcal{P}'(\mathfrak{v}^1)}{2}\right)(\mathfrak{v}^1-\mathfrak{v}^0)-\int_0^1\mathcal{P}'(\mathfrak{b}\mathfrak{v}(t))dt(\mathfrak{v}^1-\mathfrak{v}^0)\right\|_{\ell^\infty} \\ &\quad+\left\|\int_0^1\mathcal{P}'(\mathfrak{b}\mathfrak{v}(t))dt\left((\mathfrak{v}^1-\mathfrak{v}^0)-\mathfrak{b}(\mathfrak{v}^1-\mathfrak{v}^0)\right)\right\|_{\ell^\infty} \\ &\lesssim \left(h+\|\mathfrak{v}^1-\mathfrak{v}^0\|_{\ell^\infty}+\|\mathfrak{v}^1-\mathfrak{v}^0\|_{\ell^\infty}^2\right)\|\mathfrak{v}^1-\mathfrak{v}^0\|_{\ell^\infty} \\ &\lesssim h\|\mathfrak{v}^1-\mathfrak{v}^0\|_{\ell^\infty}. \end{aligned}$$

It follows that R is a retraction for h small enough and by Banach's fixpoint theorem that there exists a unique fixpoint. We now prove that this unique fixpoint $\mathfrak{v}$ of R satisfies $\mathcal{P}\mathfrak{b}\mathfrak{v}=\mathfrak{u}$. Note that we have

$$\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathfrak{u} = \mathcal{P}\mathfrak{b}\mathcal{P}\left(\mathfrak{v}-\mathfrak{b}^{-1}\mathcal{P}\mathfrak{b}\mathfrak{v}+\mathfrak{b}^{-1}\mathfrak{u}\right)-\mathfrak{u}.$$

Using the triangle inequality we have

$$\begin{aligned} \|\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathfrak{u}\|_{\ell^\infty} &\leq \left\|\mathcal{P}\mathfrak{b}\mathcal{P}\left(\mathfrak{v}-\mathfrak{b}^{-1}\mathcal{P}\mathfrak{b}\mathfrak{v}+\mathfrak{b}^{-1}\mathfrak{u}\right)-\mathcal{P}\mathfrak{b}\left(\mathfrak{v}-\mathfrak{b}^{-1}\mathcal{P}\mathfrak{b}\mathfrak{v}+\mathfrak{b}^{-1}\mathfrak{u}\right)\right\|_{\ell^\infty} \\ &\quad+\left\|\mathcal{P}\mathfrak{b}\left(\mathfrak{v}-\mathfrak{b}^{-1}\mathcal{P}\mathfrak{b}\mathfrak{v}+\mathfrak{b}^{-1}\mathfrak{u}\right)-\mathfrak{u}\right\|_{\ell^\infty} \end{aligned}$$

The first term can by Lemma B.3 be estimated by

$$\|\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathfrak{u}\|_{\ell^\infty}\left(\|\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathfrak{u}\|_{\ell^\infty}+\|\nabla\mathfrak{v}\|_{\ell^\infty}\right).$$

Using Taylor expansion and Lemma A.1 we get for the second term

$$\begin{aligned}\left\|\mathcal{P}\mathfrak{b}\left(\mathfrak{v}-\mathfrak{b}\mathcal{P}\mathfrak{b}\mathfrak{v}+\mathfrak{b}^{-1}\mathfrak{u}\right)-\mathfrak{u}\right\|_{\ell^\infty} &= \|\mathcal{P}\left(\mathfrak{b}\mathfrak{v}-\mathcal{P}\mathfrak{b}\mathfrak{v}+\mathfrak{u}\right)-\mathfrak{u}\|_{\ell^\infty}\\ &\lesssim \|\mathcal{P}\mathfrak{b}\mathfrak{v}+P_{T_{\mathcal{P}\mathfrak{b}\mathfrak{v}}M}\left(-\mathcal{P}\mathfrak{b}\mathfrak{v}+\mathfrak{u}\right)-\mathfrak{u}\|_{\ell^\infty}+\|\mathfrak{u}-\mathcal{P}\mathfrak{b}\mathfrak{v}\|_{\ell^\infty}^2\\ &= \|\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathfrak{u}-P_{T_{\mathcal{P}\mathfrak{b}\mathfrak{v}}M}\left(\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathfrak{u}\right)\|_{\ell^\infty}+\left\|\mathfrak{u}-\mathcal{P}a^{-1}\mathfrak{v}\right\|_{\ell^\infty}^2\\ &\lesssim \|\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathfrak{u}\|_{\ell^\infty}^2\,.\end{aligned}$$

Hence, we have

$$\|\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathfrak{u}\|_{\ell^\infty}\lesssim\|\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathfrak{u}\|_{\ell^\infty}\left(\|\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathfrak{u}\|_{\ell^\infty}+\|\nabla\mathfrak{v}\|_{\ell^\infty}\right). \tag{3.32}$$

As $\|\nabla\mathfrak{v}\|_{\ell^\infty}\lesssim h$ and

$$\|\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathfrak{u}\|_{\ell^\infty}\le\|\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathcal{P}\mathfrak{v}\|_{\ell^\infty}+\|\mathfrak{v}-\mathfrak{u}\|_{\ell^\infty}\lesssim\|\mathfrak{b}\mathfrak{v}-\mathfrak{v}\|_{\ell^\infty}+h\lesssim\|\nabla\mathfrak{v}\|+h\lesssim h,$$

we have

$$\|\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathfrak{u}\|_{\ell^\infty}\lesssim h\,\|\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathfrak{u}\|_{\ell^\infty}\,.$$

Hence, $\|\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathfrak{u}\|_{\ell^\infty}=0$ and $\mathcal{P}\mathfrak{b}\mathfrak{v}=\mathfrak{u}$. □

Next we show that the discrete derivatives of $\mathfrak{b}\mathfrak{v}-\mathfrak{u}$ can be bounded by powers of h.

Lemma 3.5.13. *Let $u\in C^m(\mathbb{R},M)$ with bounded derivatives, $\mathfrak{u}:=Su$ and $\mathfrak{v}\in M^{\mathbb{Z}}$ the unique solution of* (3.31)*. Then, we have*

$$\|\nabla^k(\mathfrak{b}\mathfrak{v}-\mathfrak{u})\|_{\ell^\infty}\lesssim h^{k+1}\qquad\text{for all } k\le m+1.$$

Proof. We prove the inequality by induction on k. By Lemma B.2 we have

$$\|\mathfrak{b}\mathfrak{v}-\mathfrak{u}\|_{\ell^\infty}\le\|\mathfrak{b}\mathfrak{v}-\mathfrak{b}\mathfrak{u}\|_{\ell^\infty}+\|\mathfrak{b}\mathfrak{u}-\mathfrak{u}\|_{\ell^\infty}\le\|\mathfrak{v}-\mathfrak{u}\|_{\ell^\infty}+\|\nabla\mathfrak{u}\|_{\ell^\infty}\lesssim h.$$

Hence the inequality is true for $k=0$. Assume now $k>0$. We have

$$0=\nabla^k(\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathfrak{u})=\nabla^k(\mathcal{P}\mathfrak{b}\mathfrak{v}-\mathcal{P}\mathfrak{u}). \tag{3.33}$$

Let $i\in\mathbb{Z}$. Taylor expansion of $\mathcal{P}$ at $q\in\mathbb{R}^n$ up to order k yields

$$\begin{aligned}&\mathcal{P}((\mathfrak{b}\mathfrak{v})_i)-\mathcal{P}((\mathfrak{u})_i)\\ =\;&\sum_{l=0}^{k}\frac{1}{l!}\mathcal{P}^{(l)}(q)\left[(\mathfrak{b}\mathfrak{v})_i-q,\dots,(\mathfrak{b}\mathfrak{v})_i-q\right]-\frac{1}{l!}\mathcal{P}^{(l)}(q)\left[(\mathfrak{u})_i-q,\dots,(\mathfrak{u})_i-q\right]\\ &+\mathcal{O}\left(|(\mathfrak{b}\mathfrak{v})_i-q|^{k+1}+|(\mathfrak{u})_i-q|^{k+1}\right)\\ =\;&\sum_{l=0}^{k}\frac{1}{l!}\sum_{l'=1}^{l}\mathcal{P}^{(l)}(q)\left[\underbrace{(\mathfrak{b}\mathfrak{v})_i-q,\dots,(\mathfrak{b}\mathfrak{v})_i-q}_{l'-1},(\mathfrak{b}\mathfrak{v}-\mathfrak{u})_i,\underbrace{(\mathfrak{u})_i-q,\dots,(\mathfrak{u})_i-q}_{l-l'}\right]\end{aligned}$$

$$+\mathcal{O}\left(|(\mathfrak{b}\mathfrak{v})_i - q|^{k+1} + |(\mathfrak{u})_i - q|^{k+1}\right).$$

Applying ∇^k and using Lemma B.1 yields that

$$\begin{aligned}
0 &= \left|\nabla^k \left(\mathcal{P}((\mathfrak{b}\mathfrak{v})_i) - \mathcal{P}((\mathfrak{u})_i)\right)\right| && (3.34)\\
&\lesssim \sum_{l=0}^{k}\sum_{l'=0}^{l-1}\sum_{m_1+\dots m_l=k} |\mathcal{P}|_{C^l}\left(\prod_{j=1}^{l'-1}\sum_{r=i}^{i+k}|\left(\nabla^{m_j}\left(\mathfrak{b}\mathfrak{v}-q\right)\right)_r|\right) && (3.35)\\
&\quad \sum_{r=i}^{i+k}|\left(\nabla^{m_{l'}}\left(\mathfrak{b}\mathfrak{v}-\mathfrak{u}\right)\right)_r|\left(\prod_{j=l'+1}^{l}\sum_{r=i}^{i+k}|\left(\nabla^{m_j}\left(\mathfrak{u}-q\right)\right)_r|\right) && (3.36)\\
&\quad +\sum_{r=i}^{i+k}|(\mathfrak{b}\mathfrak{v}-q)_r|^{k+1} + \sum_{r=i}^{i+k}|(\mathfrak{u}-q)_r|^{k+1}. && (3.37)
\end{aligned}$$

Choosing $q := (\mathfrak{u})_i$ the last term can be estimated by h^{k+1}. By Lemma B.2 we have

$$\begin{aligned}
|(\mathfrak{b}\mathfrak{v}-q)_r| &\lesssim |(\mathfrak{b}(\mathfrak{v}-\mathfrak{u}))_r| + |(\mathfrak{b}\mathfrak{u})_r - (\mathfrak{u})_r| + |(\mathfrak{u})_r - q|\\
&\lesssim \|\mathfrak{v}-\mathfrak{u}\|_{\ell^\infty} + \|\nabla\mathfrak{u}\|_{\ell^\infty} + |(\mathfrak{u})_r - q|\\
&\lesssim h.
\end{aligned}$$

Hence, the first term of (3.37) can also be estimated by h^{k+1}. Furthermore, if $m_j = 0$ we can estimate the corresponding factor by h. By $f \in C^m$ and the induction hypothesis we can estimate the factor corresponding to $j \neq l'$ by h^{m_j}. The first factors in the sum of (3.36) can be estimated by $h^{m_{l'}-1}$ if $m_{l'} < k$ and by h^k if $m_{l'} = k$. It follows that the terms (3.35), (3.36) and (3.37) except the one with $l = 1$ can be estimated by h^{k+1}. However, since by (3.33) the sum of all terms is zero also the term with $l = 1$ can be estimated by h^{k+1}. Hence,

$$\left|\mathcal{P}'(q)\left(\nabla^k\left(\mathfrak{b}\mathfrak{v}-\mathfrak{u}\right)\right)_i\right| = \left|\left(\mathfrak{b}\nabla^k P_{T_qM}\mathfrak{v} - \nabla^k P_{T_qM}\mathfrak{u}\right)_i\right| \lesssim h^{k+1},$$

where $P_{T_qM} = \mathcal{P}'(q)$ is the orthogonal projection onto the tangent space. Let $\tilde{\mathfrak{u}} := P_{T_qM}\mathfrak{u}$ and $\tilde{\mathfrak{v}} := P_{T_qM}\mathfrak{v}$. It follows that $\|\nabla^k\mathfrak{b}\tilde{\mathfrak{v}} - \nabla^k\tilde{\mathfrak{u}}\|_{\ell^\infty} \lesssim h^{k+1}$. Let $P_{T_qM}^{-1}\colon U \subset T_qM \to M$ be the inverse of P_{T_qM}. Note that $(P_{T_qM}^{-1})'(q) = (P'_{T_qM})^{-1}(q) = I_{T_qM}$. Taylor expansion of u at q yields

$$(\mathfrak{v})_i = q + (\tilde{\mathfrak{v}})_i - q + \sum_{l=2}^{k}(P_{T_qM}^{-1})^{(l)}(q)\left[(\tilde{\mathfrak{v}})_i - q, \dots, (\tilde{\mathfrak{v}})_i - q\right] + \mathcal{O}\left(|(\tilde{\mathfrak{v}})_i - q|^{k+1}\right)$$

By similar arguments as above one can show that

$$\|\nabla^k(\mathfrak{v}-\tilde{\mathfrak{v}})\|_{\ell^\infty} \lesssim h^{k+1} \quad \text{and} \quad \|\nabla^k(\mathfrak{u}-\tilde{\mathfrak{u}})\|_{\ell^\infty} \lesssim h^{k+1}.$$

Hence

$$\|\nabla^k\mathfrak{b}\mathfrak{v} - \nabla^k\mathfrak{u}\|_{\ell^\infty} \le \|\nabla^k\mathfrak{b}\tilde{\mathfrak{v}} - \nabla^k\tilde{\mathfrak{u}}\|_{\ell^\infty} + \|\mathfrak{b}\|_{\ell^1}\|\nabla^k(\mathfrak{v}-\tilde{\mathfrak{v}})\|_{\ell^\infty} + \|\nabla^k(\mathfrak{u}-\tilde{\mathfrak{u}})\|_{\ell^\infty} \lesssim h^{k+1}. \quad \square$$

We can now prove the main result.

Proof of Theorem 3.5.11. Using the definition of $I_{\mathcal{P}}$ (3.29), the triangle inequality, Lemma 3.5.13 and Theorem 3.5.4 we have

$$\begin{aligned}
|u - I_{\mathcal{P}}u|_{W^{l,\infty}} &= |\mathcal{P}u - \mathcal{P}Q_{\mathbb{R}^n}(\mathcal{P}\mathfrak{b})^{-1}Su|_{W^{l,\infty}} \\
&\leq |u - Q_{\mathbb{R}^n}(\mathcal{P}\mathfrak{b})^{-1}Su|_{W^{l,\infty}} \\
&\leq |u - I_{\mathbb{R}^n}u|_{W^{l,\infty}} + |I_{\mathbb{R}^n}u - Q_{\mathbb{R}^n}(\mathcal{P}\mathfrak{b})^{-1}Su|_{W^{l,\infty}} \\
&\lesssim h^{\min(m,k+1)}|u|_{W^{\min(m,k+1),\infty}} + \left|\nabla^l\left((\mathcal{P}\mathfrak{b})^{-1} - \mathfrak{b}^{-1}\right)Su\right| \\
&\lesssim h^{\min(m,k+1)}|u|_{W^{\min(m,k+1),\infty}} + \left|\nabla^l\left(\mathfrak{b}(\mathcal{P}\mathfrak{b})^{-1}Su - Su\right)\right| \\
&\lesssim h^{\min(m,k+1)}|u|_{W^{\min(m,k+1),\infty}} + h^{\min(m,k+1)+1} \\
&\lesssim h^{\min(m,k+1)}|u|_{W^{\min(m,k+1),\infty}}. \quad \square
\end{aligned}$$

3.5.4 Approximation order of the L^2 projection for nonuniform B-splines

The theory of the previous section is based on uniform B-splines. In this section we consider more generally B-splines with arbitrary knots. Then we show how to compute the L^2 projection onto the resulting space $V_{\mathcal{P}}$. We close this section by presenting some unexpected observations regarding the convergence order of the L^2 projection with nonuniform B-splines.

General B-splines

B-splines with knots $t_0 \leq \cdots \leq t_N$ are recursively defined by

$$B_{i,0}(x) := \begin{cases} 1 & t_i \leq x < t_{i+1} \\ 0 & \text{otherwise} \end{cases}, \quad \text{for all } 0 \leq i \leq N-1 \quad \text{and}$$

$$B_{i,k}(x) := \frac{x - t_i}{t_{i+k} - t_i}B_{i,k-1} + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}}B_{i+1,k-1}, \tag{3.38}$$

for all $i, k \in \mathbb{N}$ with $k + i \leq N - 1$ and $k > 0$. Note that for uniformly distributed knots, i.e. $t_i = hi$ with $h > 0$ we recover, up to a translation, the uniform B-splines of Section 3.5.1. If we have repeated knots, i.e. $t_i = t_{i+1}$ for some $i \in \{0, \ldots, N-1\}$ the denominator in the recursion formula (3.38) might be zero. In that case the term is just ignored, i.e. set to zero. For $k \in \mathbb{N}$ our finite element space is

$$\Phi := (\phi_i)_{i\in\{0,\ldots,N-k-1\}} \quad \text{where} \quad \phi_i := B_{i,k} \text{ for all } i \in \{0, \ldots, N-k-1\}.$$

The L^2 projection onto $V_{\mathcal{P}}$

The L^2 projection $Q_{L^2}u$ of a function $u \in C([0,1], M)$ onto $V_{\mathcal{P}}$ is defined by

$$Q_{L^2}u := \operatorname*{arg\,min}_{v \in V_{\mathcal{P}}} \|v - u\|_{L^2}.$$

To approximately compute it we approximate the integral with the composite midpoint rule with $R \gg N$ intervals and solve the resulting nonlinear least square problem, i.e. we have

$$Q_{L^2}u \approx \operatorname*{arg\,min}_{c=(c_j)_{j=0}^{N-k-1} \in M^{N-k-1}} \sum_{i=1}^{R} \left| \mathcal{P}\left(\sum_{j=0}^{N-k-1} c_j \phi_i\left(\frac{2i-1}{2R}\right)\right) - u\left(\frac{2i-1}{2R}\right)\right|^2.$$

We find the minimizer by an adaptation of the Gauss-Newton method (Algorithm 3). Convergence of this algorithm is an open problem. A somewhat open problem is how large the number of quadrature nodes R has to be chosen. It was observed that $R = 3N$ is already enough to observe the convergence orders stated below.

Algorithm 3 Gauss-Newton for computation of the L^2 projection onto $V_{\mathcal{P}}$

Input: $u\colon [0,1] \to M$, $N \in \mathbb{N}$, $t_0 \le \cdots \le t_N$, $R \in \mathbb{N}$ and $tol > 0$.
Output: Approximation of the nodes $c = (c_i)_{i=0}^{N-k-1} \in M^{N-k}$ of the L^2 projection of u onto $V_{\mathcal{P}}$

Choose a first guess $c^{(0)} = (c_i^{(0)})_{i=0}^{N-k-1} \in M^{N-k}$ for c, e.g. $c_i^{(0)} = u(t_i)$.
Define the matrix $A \in \mathbb{R}^{R \times N-k}$ by $A(i,j) = \phi_j\left(\frac{2i-1}{2R}\right)$ and the vector $b \in \mathbb{R}^{nR}$ by $b(i) = u\left(\frac{2i-1}{2R}\right)$.
Set $l = 0$
repeat
  Find $S \in \mathbb{R}^{n(N-k)\times \dim(M)(N-k)}$ such that $\{Sx | x \in \mathbb{R}^{\dim(M)(N-k)}\} = T_{c^{(l)}} M^{N-k}$.
  solve the linear least square problem

$$\operatorname*{arg\,min}_{x \in \mathbb{R}^{\dim(M)(N-k)}} \left| \mathcal{P}'\left((A \otimes I_n)c^{(l)}\right)(A \otimes I_n)Sx - \log_{\mathcal{P}\left((A\otimes I_n)c^{(l)}\right)}(b)\right|$$

  Update $c^{(l+1)} = \exp_{c^{(l)}}(Sx)$ or $c^{(l+1)} = e_{c^{(l)}}(Sx)$.
  $l = l + 1$
until $\max_i d(c_i^{(l)}, c_i^{(l-1)}) < tol$
return $c^{(l)}$.

Figure 3.3: Convergence order for a sphere-valued function

k	1	2	3	4	5
Uniform knots	2.0000	2.9986	3.9968	4.9764	6.0011
Open knots	2.0000	2.9984	3.4968	3.4964	3.4860
Alternating lengths	1.9999	2.9960	3.9986	2.9904	3.9893

Experiment to measure convergence rate

To numerically determine the convergence order we compute the L^2 projection using algorithm 3 for the sphere-valued function

$$u(x) = \mathcal{P}\left(\begin{pmatrix} 1 \\ x \\ \sin(x) \end{pmatrix}\right).$$

Convergence rate for uniform knot vector

By Section 3.5.4 there exists an approximation operator $Q\colon C([0,1],M) \to V_{\mathcal{P}}$ for uniform B-splines such that for $u \in H^m([0,1],M)$ we have $\|u - Qu\|_{L^2} \lesssim h^{\min(k+1,m)}\|u\|_{H^m}$ for all h sufficiently small, i.e. Q has convergence order $\min(k+1,m)$. By the definition of Q_{L^2} it follows that Q_{L^2} has convergence order at least $\min(k+1,m)$. Figure 3.5.4 suggests that the convergence order is equal to $\min(k+1,m)$.

Convergence rate for open uniform knot vectors

A problem with using uniform B-splines to solve partial differential equations is that imposing boundary conditions can be quite difficult because there are several basis functions that are nonzero at the interval ends. Hence not only coefficients c_i (as for Lagrange type basis functions) but averages of coefficients c_i have to be fixed. A solution of this problem is to use open uniform knot vectors instead, i.e. by starting with a uniform knot vector with first and last knot equal to the interval ends and then repeating the first and last knot k times. Then for both interval ends there is only one basis function which is nonzero there and imposing boundary conditions becomes easier. However, Figure 3.5.4 suggest that the convergence order is bounded by $\min(3.5, k+1)$. This is different from the linear case where we still have the optimal convergence order.

Convergence rate for knot vectors with alternating interval lenghts

For $\alpha \in (0,1) \setminus \left\{\frac{1}{2}\right\}$ we consider the knot vector with $t_0 = 0$ and

$$t_{i+1} - t_i = \begin{cases} \alpha h & \text{if } i \text{ is odd} \\ (1-\alpha)h & \text{if } i \text{ is even.} \end{cases}$$

From Figure 3.5.4 we can see that the convergence order seems to be bounded by $\min(4, k+1)$ for k odd and by $\min(3, k+1)$ for k even.

4 Variational Problems

In this chapter, we study the two numerical methods we mentioned in the introduction to minimize the harmonic energy $\mathcal{J}(u) := |u|^2_{H^1}$ (See (0.5)). Under the assumption that the boundary data are concentrated on a small enough ball it can be shown that $\mathcal{J}$ has a unique minimizer.

Theorem 4.0.1 (Jäger–Kaul [26]). *Let $\Omega \subset \mathbb{R}^s$, M a Riemannian manifold with sectional curvature bounded from above by $\kappa > 0$, $p \in M$, $r < \frac{\pi}{2\sqrt{k}}$, $B_r(p) := \{q \in M \mid d_g(p,q) < r\}$ and $\varphi\colon \partial\Omega \to B_r(a)$. Assume that for any two points in $B_r(p)$ there exists a unique geodesic connecting them. Then there exists a unique $u \in C(\Omega, B_r(p))$ with $u|_{\partial\Omega} = \varphi$ and*

$$\mathcal{J}(v) \geq \mathcal{J}(u) \text{ for all } v \in C(\Omega, M) \text{ with } v|_{\partial\Omega} = \varphi.$$

In Section 4.1 we prove that $\mathcal{J}$ behaves quadratically around this minimizer u with respect to the H^1-seminorm, i.e. we have $J(v) - J(u) \sim |v-u|^2_{H^1}$. This will be important for proving error estimates. Then, in Section 4.2 resp. 4.3, we discuss the finite distance resp. the geometric finite element method.

4.1 Ellipticity

We use the following definition of ellipticity of functionals on functions with values in a Riemannian submanifolds of $\mathbb{R}^n$. A definition which also works for a manifold not given as an embedding and a corresponding theory can be found in [24].

Definition 4.1.1. Let $M \subset \mathbb{R}^n$ be an embedded submanifold and $\Omega \subset \mathbb{R}^s$. A functional $\mathcal{J}\colon H \subset H^1(\Omega, M) \to \mathbb{R}$ is called *elliptic* around $u \in H$ if there exists $\lambda, \Lambda, \epsilon > 0$ such that for all $v \in L^\infty(\Omega, M) \cap H$ with $\|v - u\|_{L^\infty} < \epsilon$ we have

$$\lambda |v-u|^2_{H^1} \leq \mathcal{J}(v) - \mathcal{J}(u) \leq \Lambda |v-u|^2_{H^1}.$$

We now prove that the harmonic energy functional is elliptic around the unique minimizer. First we show an optimality condition for the minimizer. It is a generalization of Proposition 1.2.5.

Proposition 4.1.2. *Let M be a Riemannian submanifold of $\mathbb{R}^n$ such that the closest point projection $\mathcal{P}\colon U \subset \mathbb{R}^n \to M$ is two times differentiable with bounded derivatives, $\Omega \subset \mathbb{R}^s$, $\varphi\colon \partial\Omega \to M$, $H^1_\varphi(\Omega, M) := \{w \in H^1(\Omega, M) | w = \varphi \text{ on } \partial\Omega\}$, $\mathcal{J}\colon H^1_\varphi(\Omega, M) \to \mathbb{R}$ the harmonic energy defined by $\mathcal{J}(u) = |u|^2_{H^1}$, and $u \in H^2(\Omega, M)$ a critical point of $\mathcal{J}$. Then we have $\Delta u(x) \in T_{u(x)}M^\perp$ almost everywhere.*

Proof. Let $w \in H^1(\Omega, TM)$ be a vector field with $w(x) \in T_{u(x)}M$ for all $x \in \Omega$ and $w(x) = 0$ for all $x \in \partial\Omega$. Define $u^\epsilon \in H^1(\Omega, M)$ by $u^\epsilon(x) = \mathcal{P}(u(x) + \epsilon w(x))$ where $\mathcal{P}$ is the closest point projection. As u is a critical point of $\mathcal{J}$ we have $0 = \frac{d}{d\epsilon}\mathcal{J}(u^\epsilon)|_{\epsilon=0}$. Using Lemma 1.2.4 and integration by parts we have

$$\begin{aligned} 0 &= \frac{d}{d\epsilon}\mathcal{J}(u^\epsilon)|_{\epsilon=0} \\ &= \frac{d}{d\epsilon}\langle \nabla u^\epsilon, \nabla u^\epsilon\rangle_{L^2}|_{\epsilon=0} \\ &= 2\left\langle \frac{d}{d\epsilon}\nabla u^\epsilon|_{\epsilon=0}, \nabla u \right\rangle_{L^2} \\ &= 2\langle \nabla w, \nabla u\rangle_{L^2} \\ &= -2\langle w, \Delta u\rangle_{L^2}. \end{aligned}$$

As this equation holds for all vector fields $w \in H^1(\Omega, TM)$ with $w(x) \in T_{u(x)}M$ we have $\Delta u(x) \in T_{u(x)}M^\perp$ almost everywhere. □

Our ellipticity estimate also includes the classical Poincaré constant.

Definition 4.1.3. The *Poincaré constant* of a domain $\Omega \subset \mathbb{R}^s$ is the smallest positive number C_P such that $\|u\|_{L^2} \leq C_P|u|_{H^1}$ for all

$$u \in H^1_0 := \{u \in H^1(\Omega, \mathbb{R}^n) \mid u = 0 \text{ on } \partial\Omega\}.$$

We can now state and prove ellipticity of the harmonic energy.

Proposition 4.1.4. *Let M be a Riemannian submanifold of $\mathbb{R}^n$ such that the closest point projection $\mathcal{P}\colon U \subset \mathbb{R}^n \to M$ is two times differentiable with bounded derivatives, $\Omega \subset \mathbb{R}^s$, $\varphi\colon \partial\Omega \to M$, $H^1_\varphi(\Omega, M) := \{w \in H^1(\Omega, M) | w = \varphi \text{ on } \partial\Omega\}$, $\mathcal{J}\colon H^1_\varphi(\Omega, M) \to \mathbb{R}$ the harmonic energy $\mathcal{J}(u) = |u|^2_{H^1}$ and $u \in W^{2,\infty}(\Omega, M)$ a critical point of $\mathcal{J}$. Then if $\|\Delta u\|_{L^\infty} < C_P^{-2}|\mathcal{P}|^{-1}_{C^2}$ where C_P is the Poincaré constant from Definition 4.1.3, the function $\mathcal{J}$ is elliptic around u.*

Proof. For $v \in H^1_\varphi(\Omega, M)$ we have

$$\begin{aligned} \mathcal{J}(v) - \mathcal{J}(u) &= \langle \nabla v - \nabla u, \nabla v + \nabla u\rangle_{L^2} \\ &= \langle \nabla v - \nabla u, \nabla v - \nabla u\rangle_{L^2} + 2\langle \nabla v - \nabla u, \nabla u\rangle_{L^2} \\ &= |v - u|^2_{H^1} - 2\langle v - u, \Delta u\rangle_{L^2}. \end{aligned}$$

We define the function $z \in H^1(\Omega, TM)$ by $z(x) := P_{T_{u(x)}M}[v(x) - u(x)]$ for all $x \in \Omega$. Since $z(x) \in T_{u(x)}M$ and $\Delta u(x) \in T_{u(x)}M^\perp$ (Proposition 4.1.2) we have $\langle z, \Delta u\rangle_{L^2} = 0$. For $\|v-u\|_{L^\infty}$ small enough we can define the function $\gamma(x,t) := \mathcal{P}(tv(x) + (1-t)u(x))$. By Lemma 1.2.4 we have $\mathcal{P}'(u(x)) = P_{T_{u(x)}M}$ and hence

$$\begin{aligned} |v(x) - u(x) - z(x)| &= |\gamma(x,1) - \gamma(x,0) - \dot\gamma(x,0)| \\ &= \left|\int_0^1 (1-t)\ddot\gamma(x,t)dt\right| \\ &\leq \frac{1}{2}|\mathcal{P}|_{C^2}|v(x) - u(x)|^2. \end{aligned}$$

By the Poincaré inequality we get

$$\begin{aligned} |\langle v-u, \Delta u\rangle_{L^2}| &= |\langle z, \Delta u\rangle_{L^2} - \langle v-u-z, \Delta u\rangle_{L^2}| \\ &= |\langle v-u-z, \Delta u\rangle_{L^2}| \\ &\leq \|v-u-z\|_{L^1}\|\Delta u\|_{L^\infty} \\ &\leq \frac{1}{2}|\mathcal{P}|_{C^2}\|v-u\|_{L^2}^2\|\Delta u\|_{L^\infty} \\ &\leq \frac{1}{2}|\mathcal{P}|_{C^2}C_P^2|v-u|_{H^1}^2\|\Delta u\|_{L^\infty}. \end{aligned}$$

It follows that

$$(1 - |\mathcal{P}|_{C^2}C_P^2\|\Delta u\|_{L^\infty})|v-u|_{H^1}^2 \leq \mathcal{J}(v) - \mathcal{J}(u) \leq (1 + |\mathcal{P}|_{C^2}C_P^2\|\Delta u\|_{L^\infty})|v-u|_{H^1}^2.$$

For $\|\Delta u\|_{L^\infty} < C_P^{-2}|\mathcal{P}|_{C^2}^{-1}$ we get the desired estimate. □

For a half-sphere HS^n we can estimate $|\mathcal{P}|_{C^2}$ by Lemma A.3 and get the following Corollary

Corollary 4.1.5. *Let $\varphi\colon \Omega \to HS^n$, $H^1_\varphi(\Omega, HS^n)$ and $\mathcal{J}$ as in Proposition 4.1.4 and $u \in W^{2,\infty}([0,1]^s, HS^n)$ a critical point of $\mathcal{J}$. Then if*

$$\|\Delta u\|_{L^\infty} < \frac{\sqrt{3}}{2C_P^2},$$

the functional $\mathcal{J}$ is elliptic around u.

4.2 The finite distance method

In this section we construct a simple method to minimize the harmonic energy on the hypercube $\Omega = [0,1]^s$ with given boundary data $g\colon \partial\Omega \to M$. Consider first the

hypercube $[0,h]^s$ for $h > 0$. We approximate the harmonic energy $\mathcal{J}$ of a function $u \in H^1([0,h]^s, \mathbb{R}^n)$ by

$$\tilde{\mathcal{J}}(u) := \frac{h^{s-2}}{2^{s-1}} \sum_{(i,j)\in E([0,h]^s)} d^2(u(i), u(j)), \tag{4.1}$$

where $E([0,h]^s)$ denotes the set of all edges of the hypercube $[0,h]^s$ and $d\colon M \times M \to \mathbb{R}_{\geq 0}$ is a distance with $d(u,v) = |v-u| + \mathcal{O}(|v-u|^3)$. By Lemma C.1 and the classical error estimates of the trapezoidal rule there exists $C > 0$ such that we have

$$\left|\mathcal{J}(u) - \tilde{\mathcal{J}}(u)\right| \leq Ch^{s+2}\|u\|^2_{C^3} \text{ for all } u \in C^3([0,h]^s, \mathbb{R}^n). \tag{4.2}$$

Let us now for $N \in \mathbb{N}$ subdivide our cube $[0,1]^s$ into N^s smaller cubes, each of side length N^{-1}. Consider also the corresponding vertices $V := \{0, N^{-1}, 2N^{-1}, \dots, (N-1)N^{-1}, 1\}^s$, and the edge set $E := \{(i,j) \in V \times V \mid |i-j| = N^{-1}\}$. Using the discretization (4.1) and adding up the results yields the approximation

$$\mathcal{J}_N(u) := N^{2-s} \sum_{(i,j)\in E} w_{(i,j)} d^2(u(i), u(j)),$$

where the weight $w_{(i,j)}$ is the number of small cubes which have $(i,j) \in E$ as an edge divided by 2^{s-1}. For an interior edge this weight is 1 while for an edge lying on the boundary the weight is smaller than 1. Summing up all error estimates (4.2) from the small cubes yields that there exists $C > 0$ such that we have

$$|\mathcal{J}(u) - \mathcal{J}_N(u)| \leq CN^{-2}\|u\|^2_{C^3} \text{ for all } u \in C^3([0,1]^s, \mathbb{R}^n). \tag{4.3}$$

Note that $\mathcal{J}_N(u)$ depends only on the values of u on V. Furthermore the values of u at the boundary vertices $V_B \subset V$ are given by $\varphi|_{V_B}\colon V_B \to \mathbb{R}^n$. Let $J_N\colon (HS^n)^{V_I} \to \mathbb{R}$ be the functional $\mathcal{J}_N$ restricted to the values at the interior vertices $V_I := V \backslash V_B$, i.e.

$$J_N(u) := N^{2-s} \left(\sum_{\substack{(i,j)\in E,\\ i\in V_B,\\ j\in V_B}} w_{(i,j)} d^2\left(\varphi(i), \varphi(j)\right) + \sum_{\substack{(i,j)\in E,\\ i\in V_B,\\ j\in V_I}} d^2\left(\varphi(i), u(j)\right) + \sum_{\substack{(i,j)\in E,\\ i\in V_I,\\ j\in V_I}} d^2\left(u(i), u(j)\right) \right). \tag{4.4}$$

By Proposition 2.3.1 and Theorem 2.4.2 this functional has a unique minimizer if M is a Hadamard manifold or the open half-sphere.

Note that if $u \in M^V$ is a minimizer of $J_N(u)$ we have for all $i \in V$ that u_i is the Riemannian (resp. projection) average of the value of its neighbors, i.e. we have a discrete mean-value property.

In the derivation of the finite distance method we first discretized and then optimized. Let us try to reverse the order, i.e. to first optimize and then discretize. By Proposition 4.1.2 the optimality condition in the case where M is a Riemannian submanifold of $\mathbb{R}^n$ is $\Delta u(x) \in T_{u(x)}M^\perp$ almost everywhere. Discretization of this condition on a regular grid in a natural way would yield again that the value at some point on the grid is the average of the values of its neighbors.

4.2.1 Algorithms to minimize the discrete harmonic energy

In this section we study some algorithms to minimize J_N. The mean-value property mentioned in the previous section motivates Algorithm 4 where we iteratively replace the values by the Riemannian or projection-based average of its neighbors.

Algorithm 4 Iterative averaging for minimization of the discrete harmonic energy

Input: Boundary data $g\colon \partial\Omega^2 \to M$ where $\Omega = [0,1]^2$ and parameters $N \in \mathbb{N}$ and $tol > 0$.

Output: Approximation of a minimizer of the harmonic energy on the regular $(N+1)\times(N+1)$ grid of the unit square.

Choose a first guess $(u^{(0)})_{i,j=0}^{N}$ for u (e.g. by interpolation of g)
Set $k = 0$
repeat
  for $i = 1 : N-1$ **do**
    for $j = 1 : N-1$ **do**
      $u_{i,j}^{(k+1)} = av\left((u_{i-1,j}^{(k)}, u_{i+1,j}^{(k)}, u_{i,j-1}^{(k)}, u_{i,j+1}^{(k)}), \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)\right)$
    end for
  end for
  $k = k + 1$
until $\max_{i,j} d(u_{i,j}^{(k)}, u_{i,j}^{(k-1)}) < tol$
return $u^{(k)}$.

To get a second order method one can use Algorithm 5 which makes use of the Riemannian Newton method introduced in Section 1.3.3.

4.2.2 Convergence analysis

Given the discrete minimizer of J_N we construct in this section an approximation for the minimizer of $\mathcal{J}$. Then we prove that this approximation converges to the minimizer of $\mathcal{J}$ for $N \to \infty$.

Interpolation of the values on a grid

To construct an approximation we define finite element functions and then use the approximation operator $Q_{\mathcal{P}}$ from Section 3.3. For $i = (i_1, \ldots, i_d) \in \{0, 1, \ldots, N\}^s$ and $x = (x_1, \ldots, x_d) \in [0,1]^s$ we define the tensor product basis function $\phi_i\colon [0,1]^s \to \mathbb{R}$ by

$$\phi_i(x) := \prod_{j=1}^{s} t_{i_j}(x_j) \in \mathbb{R},$$

where for $k \in \{0, 1, \ldots, N\}$ the function $t_k\colon [0,1] \to \mathbb{R}$ is the piecewise linear function with $t_k(lN^{-1}) = \delta_{kl}$.

Algorithm 5 Minimization of the discrete harmonic energy with the Newton method

Input: Boundary data $g\colon \partial\Omega^2 \to M$ where $\Omega = [0,1]^2$ and parameters $N \in \mathbb{N}$ and $tol > 0$.

Output: Approximation of a minimizer of the harmonic energy on the regular $(N+1)\times(N+1)$ grid of the unit square.

Choose a first guess $(u^{(0)})_{i,j=0}^{N}$ for u (e.g. by interpolation of g)
Set $k = 0$
repeat
Compute gradient and Hessian of

$$J(u) = \sum_{i=0}^{N}\sum_{j=0}^{N-1} d^2(u_{i,j}, u_{i,j+1}) + \sum_{i=0}^{N-1}\sum_{j=0}^{N} d^2(u_{i,j}, u_{i+1,j})$$

at $u^{(k)}$
Restrict Hess and gradient grad to the interior points.
Solve $\operatorname{Hess} J(u^{(k)})x = \operatorname{grad} J(u^{(k)})$ for $x \in T_{u^{(k)}}M^{(N-1)\times(N-1)}$
Update $u^{(k+1)} = \exp_{u^{(k)}}(-x)$ or $u^{(k+1)} = e_{u^{(k)}}(-x)$
$k = k+1$
until $\max_{i,j} d(u_{i,j}^{(k)}, u_{i,j}^{(k-1)}) < tol$
return $u^{(k)}$.

Convergence for the half-sphere case

We now show that if M is the half-sphere HS^n the constructed approximation converges towards the unique minimizer of $\mathcal{J}$. However, we need to assume that the distance $d\colon M \times M \to \mathbb{R}_{\geq 0}$ is of the form $d^2(p,q) = \alpha(\langle p, q\rangle)$ with α convex, monotonically decreasing and $\alpha(x) \geq 4(1-x)(1+x)^{-1}$. The reason is that under this condition, we can show that the harmonic energy of the interpolation can be bounded by the discrete harmonic energy.

Lemma 4.2.1. *Assume that $d^2(p,q) = \alpha(\langle p, q\rangle)$ with α convex monotonically decreasing and $\alpha(x) \geq 4(1-x)(1+x)^{-1}$. Then we have*

$$\mathcal{J}(Q_{\mathcal{P}}u) \leq J_N(u)$$

for all $u \in (HS^n)^V$.

Proof. Note that it is enough to prove the estimate for $N = 1$. For $q \in \mathbb{R}^{n+1}\backslash\{0\}$ and $r \in \mathbb{R}^{n+1}$ we have

$$\mathcal{P}'(q)[r] = \frac{r}{|q|} - \frac{\langle r, q\rangle q}{|q|^3}.$$

Hence,

$$\left|\mathcal{P}'(q)[r]\right|^2 = \frac{|r|^2}{|q|^2} - \frac{\langle r, q\rangle^2}{|q|^4} \leq \frac{|r|^2}{|q|^2}.$$

By Lemma C.2, the Cauchy–Schwarz and Jensen's inequality we have for $x \in [0,1]^s$ with $\phi_{i+j} = \phi_i + \phi_j$

$$\begin{aligned}
\sum_{k=1}^{s} \left| \partial_k \mathcal{P}\left(\sum_{i \in V} \phi_i(x) u_i \right) \right|^2 &= \sum_{k=1}^{s} \left| \mathcal{P}'\left(\sum_{i \in V} \phi_i(x) u_i \right) \left[\sum_{i \in V} \partial_k \phi_i(x) u_i \right] \right|^2 \\
&\leq \frac{\sum_{(i,j)\in E} \phi_{i,j}(x) |u_i - u_j|^2}{1 - \frac{1}{4} \sum_{(i,j)\in E} \phi_{i,j} |u_i - u_j|^2} \\
&\leq \sum_{(i,j)\in E} \phi_{i,j}(x) \frac{|u_i - u_j|^2}{1 - \frac{1}{4}|u_i - u_j|^2} \\
&= \sum_{(i,j)\in E} \phi_{i,j}(x) 4(1 - \langle u_i, u_j \rangle)(1 + \langle u_i, u_j \rangle)^{-1} \\
&\leq \sum_{(i,j)\in E} \phi_{i,j}(x) \alpha(\langle u_i, u_j \rangle) \\
&= \sum_{(i,j)\in E} \phi_{i,j}(x) d^2(u_i, u_j).
\end{aligned}$$

Hence, we have

$$\begin{aligned}
\mathcal{J}(Q_{\mathcal{P}} u) &= \int_{[0,1]^s} \sum_{k=1}^{s} \left| \partial_k P\left(\sum_{i \in V} \phi_i(x) u_i \right) \right|^2 dx \\
&\leq \int_{[0,1]^s} \sum_{(i,j)\in E} \phi_{i,j}(x) d^2(u_i, u_j) dx = \frac{1}{2^{s-1}} \sum_{(i,j)\in E} d^2(u_i, u_j) = J_N(u). \qquad \square
\end{aligned}$$

We first show convergence in the L^∞-norm.

Theorem 4.2.2. *Let $\varphi \in C(\partial[0,1]^s, HS^n)$, for every $N \in \mathbb{N}$, J_N the functional defined in (4.4) with α as in Lemma 4.2.1, $u_N \in (HS^n)^{V_I}$ the unique minimizer of J_N and $u \in H^1([0,1]^s, S^n)$ the minimizer of $\mathcal{J}$. Assume that $u \in C^3$. Then we have*

$$\lim_{N \to \infty} \| Q_{\mathcal{P}} u_N - u \|_{L^\infty} = 0$$

Proof. Assume that the statement is not true. Then there exists $\epsilon > 0$ and a subsequence $(N_i)_{i \in \mathbb{N}} \subset \mathbb{N}$ with

$$\| Q_{\mathcal{P}} u_{N_i} - u \|_{L^\infty} \geq \epsilon \text{ for all } i \in \mathbb{N}.$$

As u_N is the minimizer of J_N we have

$$J_N(u_N) - \mathcal{J}_N(u) \leq 0. \tag{4.5}$$

By Lemma 4.2.1, (4.5) and Inequality (4.3) there exists $C > 0$ such that we have

$$\mathcal{J}(Q_{\mathcal{P}} u_N) \leq J_N(u_N) \leq \mathcal{J}_N(u) \leq \mathcal{J}(u) + CN^{-2}.$$

Hence, $\mathcal{J}(Q_{\mathcal{P}}u_{N_i}) = |Q_{\mathcal{P}}u_{N_i}|_{H^1}$ is bounded. It follows that $(Q_{\mathcal{P}}u_{N_i})_{i\in\mathbb{N}}$ has a subsequence which convergences weakly to a function $v \in H^1([0,1]^s, \mathbb{R}^{n+1})$. Since Dirac measures are linear functionals on L^∞ it follows that $v \in H^1([0,1]^s, S^n)$. By weakly lower semicontinuity of the H^1-norm we have $|v|_{H^1} \leq |u|_{H^1}$, which is a contradiction to the uniqueness of the minimizer of $\mathcal{J}$ by Theorem 4.0.1. □

The restriction of the function $\mathcal{P}Iu_N$ to the boundary is $\mathcal{P}I\varphi_N$ and in general not equal to the boundary data φ of the function u. We denote by $\tilde{u}_N$ the unique minimizer with respect to the boundary data $\mathcal{P}I\varphi_N$. We will need to estimate $\|u - \tilde{u}_N\|_{H^1(\Omega,\mathbb{R}^n)}$. In the linear case this can be done by linearity and [27]. In the manifold-valued case this is an open problem.

Conjecture 4.2.3. *Let M be a Riemannian manifold, $\Omega \subset \mathbb{R}^s$ open and bounded with smooth boundary, $\varphi \in C(\partial\Omega, HS^n)$, $\tilde{u}_N$ the unique minimizer with respect to the boundary data $\mathcal{P}I\varphi_N$. Then we have*

$$\|u - \tilde{u}_N\|_{H^1(\Omega,\mathbb{R}^n)} \lesssim \|\varphi - \mathcal{P}I\varphi_N\|_{H^{\frac{1}{2}}(\partial\Omega,\mathbb{R}^n)}. \tag{4.6}$$

Assuming Conjecture 4.2.3 holds, we can estimate the H^1-error between the exact and the approximate solution.

Theorem 4.2.4. *Let $\Omega = [0,1]^s$, $\varphi \in C(\partial\Omega, HS^n)$, $H^1_\varphi(\Omega, HS^n)$ as in Theorem 4.2.2,*

$$u := \operatorname*{arg\,min}_{v\in H^1_\varphi(\Omega,HS^n)} \mathcal{J}(v),$$

J_N the discrete harmonic energy and $u_N \in (HS^n)^{V_{Int}}$ the unique minimizer of J_N. Assume that $u \in C^3$ and $\|\Delta u\|_{L^\infty} < \frac{\sqrt{3}s\pi^2}{2}$. Then if Conjectures 4.2.3 holds we have

$$|\mathcal{P}Iu_N - u|_{H^1} \lesssim N^{-1}.$$

Proof. By Theorem 4.2.2 we have convergence in L^∞. As u_N is the minimizer of J_N we have

$$J_N(u_N) - \mathcal{J}_N(u) \leq 0. \tag{4.7}$$

Let $\tilde{u}_N \in H^1(\Omega, HS^n)$ be the unique minimizer with respect to the (interpolated) boundary data $\varphi_N = \mathcal{P}Iu_N|_{\partial\Omega}$. Using the triangle inequality we have

$$|\mathcal{P}Iu_N - u|_{H^1(\Omega,\mathbb{R}^n)} \leq |\mathcal{P}Iu_N - \tilde{u}|_{H^1(\Omega,\mathbb{R}^n)} + |\tilde{u} - u|_{H^1(\Omega,\mathbb{R}^n)}.$$

By (4.7), Lemma 4.1.4 and Inequality (4.3) we have that

$$\begin{aligned} |\mathcal{P}Iu_N - \tilde{u}|^2_{H^1(\Omega,\mathbb{R}^n)} &\lesssim \mathcal{J}(\mathcal{P}Iu_N) - \mathcal{J}(u) \\ &= (\mathcal{J}(\mathcal{P}Iu_N) - J_N(u_N)) + (J_N(u_N) - \mathcal{J}_N(u)) + (\mathcal{J}_N(u) - \mathcal{J}(u)) \\ &\leq |\mathcal{J}_N(u) - \mathcal{J}(u)| \end{aligned}$$

$$\lesssim \quad N^{-2}.$$

Using Conjecture 4.2.3, the Gagliardo Nirenberg inequality for fractional Sobolev spaces [10] and Theorem 3.3.4 we get

$$\begin{aligned} |\tilde{u} - u|_{H^1(\Omega,\mathbb{R}^n)} &\lesssim \|\varphi - \mathcal{P}I\varphi_N\|_{H^{\frac{1}{2}}(\partial\Omega,\mathbb{R}^n)} \\ &\lesssim \|\varphi - \mathcal{P}I\varphi_N\|^{\frac{1}{2}}_{H^1}(\partial\Omega,\mathbb{R}^n)\|\varphi - \mathcal{P}I\varphi_N\|^{\frac{1}{2}}_{L^2}(\partial\Omega,\mathbb{R}^n) \\ &\lesssim N^{-1}. \end{aligned}$$

Combining the inequalities yields the desired result. □

An interesting open problem is to estimate the L^2-error. A possible approach would be to generalize the Aubin–Nitsche-duality argument [42].

4.3 The geometric finite element method

In this section, we study the geometric finite element method. In Section 4.3.2, we give a convergence theory for geometric finite elements. Finally, in Section 4.3.2, we discuss some issues which occur when implementing the geometric finite element method.

4.3.1 Convergence analysis for the geometric finite element method

Having ellipticity it is straight forward to prove convergence for geometric finite elements. We first show a nonlinear Céa lemma.

Lemma 4.3.1. *Let $M \subset \mathbb{R}^n$ be a Riemannian submanifold of $\mathbb{R}^n$, $\Omega \subset \mathbb{R}^s$, $\mathcal{J}\colon H \subset H^1(\Omega, M) \to \mathbb{R}$ a functional with a unique minimizer $u \in H$. Assume that $\mathcal{J}$ is elliptic around u. Let $V \subset H$ be an nonempty subset and*

$$v := \operatorname*{arg\,min}_{w \in V} \mathcal{J}(w).$$

Then we have

$$|v - u|_{H^1} \lesssim \inf_{w \in V} |w - u|_{H^1}.$$

Proof. By the ellipticity we have for any $w \in V$

$$|v - u|^2_{H^1} \leq \mathcal{J}(v) - \mathcal{J}(u) \leq \mathcal{J}(w) - \mathcal{J}(u) \lesssim |w - u|^2_{H^1}.$$

Taking the square root yields the desired result. □

We can now prove the convergence estimate for geometric finite elements.

Theorem 4.3.2. *Let $M \subset \mathbb{R}^n$ be an embedded submanifold, $\Omega \subset \mathbb{R}^s$, $\mathcal{J}\colon H \subset H^1(\Omega, M) \to \mathbb{R}$ a functional with a unique minimizer $u \in H \cap H^m$. Assume that $\mathcal{J}$ is elliptic around u. Let $\phi = (\phi_i)_{i\in I} \subset W^{1,q}(\Omega, \mathbb{R})$ for some $q > \max(d, 2)$, $Q_{\mathbb{R}^n}$, V_R, Q_R, $V_\mathcal{P}$ and $Q_\mathcal{P}$ as in Chapter 3 and*

$$v_R := \operatorname*{arg\,min}_{w\in V_R\cap H} \mathcal{J}(w) \quad \text{and} \quad v_\mathcal{P} := \operatorname*{arg\,min}_{w\in V_\mathcal{P}\cap H} \mathcal{J}(w).$$

Assume that $Q_{\mathbb{R}^n}$ satisfies (3.4) *with $l = 1$ and $p = 2$. Then there exist a constant C_R depending only on u, M and the implicit constant of Lemma 4.3.1 and a constant $C_\mathcal{P}$ depending only on the implicit constant of Lemma 4.3.1 and Inequality* (3.4) *such that for h small enough we have*

$$|v_R - u|_{H^1} \le h^{m-1} C_R, \qquad \text{and } |v_\mathcal{P} - u|_{H^1} \le C_\mathcal{P} h^{m-1} |u|_{H^m}.$$

Proof. By Lemma 4.3.1 and Theorem 3.4.3 we have

$$|v_R - u|^2_{H^1} \lesssim |Q_R u - u|_{H^1} \lesssim h^{m-l} C_R.$$

The proof for the projection-based solution uses Theorem 3.3.4 and is analogous. □

4.3.2 Implementing the geometric finite element method

The geometric finite element method was implemented in Matlab. However as the code is in general not very efficient we omit a detailed description and concentrate ourselves to a few important tricks necessary to implement the geometric finite element method. For the projection average based finite element method we need to compute

$$\operatorname*{arg\,min}_{(c_i)_{i\in I}\subset M} \int_\Omega \sum_{j=1}^{n} \left| \frac{\partial}{\partial x_j} \mathcal{P}\left(\sum_{i\in I} \phi_i(x) c_i \right) \right|^2 dx.$$

In a first step the integral is approximated by a quadrature rule. The resulting expression can be minimized using the Riemannian Newton method. To compute the derivatives one can first compute the derivatives of the linear combination $\sum_{i\in I} \phi_i c_i$ and the closest point projection $\mathcal{P}$ at $\sum_{i\in I} \phi_i c_i$ and then use the chain rule to get the derivatives of the composition. When M is a compact Stiefel manifold we have for example

$$\mathcal{P}\left(\sum_{i\in I} \phi_i(x) c_i \right) = \left(\sum_{i\in I} \phi_i(x) c_i \right) \left(\sum_{i,j\in I} \phi_i(x)\phi_j(x) c_i c_j^T \right)^{-\frac{1}{2}}.$$

To compute derivatives of this expression one can first compute derivatives of the individual functions and then combine them using the product and the chain rule.

Higher order Lagrange basis on triangles

We briefly point out the construction of an order $p \in \mathbb{N}$ Lagrange basis on a triangular element using barycentric coordinates $(\lambda_i)_{i=1}^3$. Consider the index set $I = \{(i,j,k) \,|\, i,j,k \in \mathbb{N}, i+j+k = p\}$. For $x \in \mathbb{R}$ and $n \in \mathbb{N}$ we define the polynomial

$$\binom{x}{n} := \frac{1}{n!} \prod_{i=0}^{n-1} (x-i).$$

Note that for $v = (a,b,c) \in I$ the function

$$\phi_v(\lambda_1, \lambda_2, \lambda_3) := \binom{p\lambda_1}{a}\binom{p\lambda_2}{b}\binom{p\lambda_3}{c},$$

is the unique polynomial of (total) degree at most p with $\phi_v(w/p) = \delta_{v,w}$ for all $w \in I$.

Harmonic energy of a geometric finite element function on a triangular element

In the previous section we expressed the basis function in terms of barycentric coordinates. These expression also make sense for $\lambda_1, \lambda_2, \lambda_3$ whose sum is not equal to 1. There is also a natural way to define the projection (Definition 1.4.1) resp. Riemannian (Definition 1.4.3) average also for weights with sum not equal to 1. Hence a derivative $du/d\lambda_i$ of a geometric finite element function $u \colon T \to M \subset \mathbb{R}^n$ on a triangle $T \subset \mathbb{R}^2$ makes sense. In this section we explain how to express $\left|\frac{\partial u(x)}{\partial x_j}\right|^2$ in terms of the derivatives with respect to the barycentric coordinates $(\lambda_i)_{i=1}^3$. Note that the coordinates $x = (x_1, x_2) \in \mathbb{R}^2$ of a point inside the triangle with vertices $P_1, P_2, P_3 \in \mathbb{R}^2$ can be expressed in the variables $(\lambda_i)_{i=1}^3$ by

$$x = P_1\lambda_1 + P_2\lambda_2 + P_3\lambda_3. \tag{4.8}$$

It follows that

$$\begin{pmatrix} P_1 & P_2 & P_3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} x \\ 1 \end{pmatrix}. \tag{4.9}$$

Let $A = [a_{ij}]$ be the inverse of the 3×3 matrix

$$\begin{pmatrix} P_1 & P_2 & P_3 \\ 1 & 1 & 1 \end{pmatrix}.$$

We can now find an expression of the norm of the gradient of $u = (u_1, \ldots, u_n)$ in terms of the derivatives in the barycentric coordinates. By the chain rule we have

$$|\nabla u|^2 = \sum_{i=1}^{n}\sum_{j=1}^{2}\left(\frac{du_i}{dx_j}\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{2}\left(\sum_{k=1}^{3} a_{kj}\frac{du_i}{d\lambda_k}\right)^2 = \sum_{i=1}^{n}\sum_{k=1}^{3}\sum_{l=1}^{3}\sum_{j=1}^{2} a_{k,j}a_{l,j}\frac{du_i}{d\lambda_k}\frac{du_i}{d\lambda_l}.$$

5 Discussion

Let us take a look back to the results of this thesis and discuss some possible direction for future research in this area.

In Chapter 1, we presented a theory for the minimization of functionals and the computation of averages on Riemannian manifolds which requires only elementary knowledge of differential geometry. An important tool was the closest point projection $\mathcal{P}$. We presented a simple relation between the classical gradient and Hessian and the Riemannian gradient and Hessian (Proposition 1.3.1). We showed how this formula can for example be applied to the compact Stiefel manifold (Equation (1.23)) which includes the sphere and the special orthogonal group.

In Chapter 2, we proposed the IRM algorithm to minimize the TV functional. We presented various theories concerning the existence of unique minimizers of functionals related to the TV-functional and the convergence and convergence speed of IRM.

In Chapter 3, we showed that the projection based approximation operator $Q_{\mathcal{P}}$ satisfies the same error estimate as its linear analog $Q_{\mathbb{R}^n}$. Furthermore we showed that with uniform B-splines we have the same approximation order whereas for nonuniform B-splines we observed that the convergence order breaks down.

In Chapter 4, we presented two techniques to solve variational problems. We gave a convergence theory showing that for elliptic functionals we have optimal convergence order.

The thesis can be summarized by saying that we generalized methods and theories from the real-valued case to the manifold-valued case.

There are still many open problems (e.g. why the symmetrization in Section 1.5.3 is not necessary, uniqueness of minimizer of TV-functional with sphere-valued data (Section 2.4), convergence of Algorithm 3 and explanation of the convergence orders in Section 3.5.4). In principle any statement in the linear theory can be generalized to a more general statement for the manifold-valued case. A few examples of statements which could be generalized are Jensen's inequality, saturation theorems or error estimates for the p-method. However, the generalization is not always true (see e.g. Section 3.5.4). Furthermore, the proof from the linear theory can not always be transferred into a proof for the manifold-valued case. Sometimes new ideas are required to generalize to the manifold-valued case.

Appendices

A Estimates related to the closest point projection

The following lemma estimates the difference between p and q and its projection onto the tangent space at p.

Lemma A.1. *For $p, q \in M$ with $|q-p|$ small enough we have*

$$|q - p - P_{T_pM}(q-p)| \leq \frac{1}{2}|\mathcal{P}|_{C^2}|q-p|^2,$$

where P_{T_pM} denotes the orthogonal projection onto T_pM.

Proof. By Lemma 1.2.4 we have $P_{T_pM} = \mathcal{P}'(p)$. Let $\gamma(t) := \mathcal{P}(tq + (1-t)p)$. Note that $\ddot{\gamma}(t) = \mathcal{P}''[q-p, q-p]$. Hence we have

$$|(q-p) - \mathcal{P}'(p)[q-p]| = |\gamma(1) - \gamma(0) - \dot{\gamma}(0)| = \left|\int_0^1 (1-t)\ddot{\gamma}(t)\right| \leq \frac{1}{2}|\mathcal{P}|_{C^2}|q-p|^2. \quad \square$$

The next lemma makes a similar statement but with $q-p$ replaced by $\log_p(q)$.

Lemma A.2. *For $p, q \in M$ with $|q-p|$ small enough we have*

$$|log_p(q) - \mathcal{P}'(p)(q-p)| \lesssim |q-p|^3,$$

with implicit constant depending only on $|\mathcal{P}|_{C^2}$ and $|\mathcal{P}|_{C^3}$.

Proof. Let $\gamma \in C^1([0,1], M)$ be the geodesic with $\gamma(0) = p$ and $\gamma(1) = q$. By Lemma 1.2.4 we have that $\mathcal{P}'(p)$ is the orthogonal projection onto T_pM. Furthermore we know that $\ddot{\gamma}(0) \in T_pM^{\perp}$. Hence we have $\mathcal{P}'(p)[\ddot{\gamma}(0)] = 0$. From Proposition 1.2.6 it follows that $|\dddot{\gamma}(t)| \lesssim |q-p|^3$. Therefore we have

$$\begin{aligned}
|\log_p(q) - \mathcal{P}'(p)[q-p]| &= |\dot{\gamma}(0) - \mathcal{P}'(p)[\gamma(1) - \gamma(0)]| \\
&= |\dot{\gamma}(0) - \mathcal{P}'(p)[\dot{\gamma}(0) + \frac{1}{2}\ddot{\gamma}(0) + \mathcal{O}(\max_{t\in[0,1]} |\dddot{\gamma}(t)|^2)]| \\
&= |\mathcal{P}'(p)[\mathcal{O}(|q-p|^3)] \\
&\lesssim |q-p|^3. \quad \square
\end{aligned}$$

For Proposition 4.1.5 we need an explicit bound on the second derivative of the closest point projection onto the sphere. The next lemma gives the exact value of the norm.

Lemma A.3. *Let $\mathcal{P}$ be the closest point projection onto the sphere S^n, i.e. $\mathcal{P}(u) = u/|u|$ for all $u \in \mathbb{R}^{n+1}\backslash\{0\}$. For $u \in \mathbb{R}^{n+1}\backslash\{0\}$ we have*

$$\|\mathcal{P}''(u)\| = \frac{2}{\sqrt{3}|u|^2},$$

where $\|\cdot\|$ denotes the operator norm.

Proof. Some calculus yields

$$\frac{d^2\mathcal{P}(u)_i}{du_j du_k} = -\frac{\delta_{ij}u_k + \delta_{ik}u_j + \delta_{jk}u_i}{|u|^3} + 3\frac{u_i u_j u_k}{|u|^5}.$$

Hence for $v, w \in \mathbb{R}^{n+1}$ we have

$$\begin{aligned}
|\mathcal{P}''(u)[v,w]|^2 &= \left|\frac{-v\langle u,w\rangle - w\langle u,v\rangle - u\langle v,w\rangle}{|u|^3} + 3\frac{\langle u,v\rangle\langle u,w\rangle u}{|u|^5}\right|^2 \\
&= |u|^{-6}\left(|v|^2\langle u,w\rangle^2 + |w|^2\langle u,v\rangle^2 + |u|^2\langle v,w\rangle^2 - 3\frac{\langle u,v\rangle^2\langle u,w\rangle^2}{|u|^2}\right) \\
&= |u|^{-4}|v|^2|w|^2\left(\cos^2(\alpha) + \cos^2(\beta) + \cos^2(\gamma) - 3\cos^2(\alpha)\cos^2(\beta)\right),
\end{aligned}$$

where α, β resp. γ denote the angle between u and w, u and v resp. v and w. Without loss of generality we can choose $\alpha, \beta, \gamma \in [0, \pi)$. By the triangle inequality we have $\gamma \geq |\alpha - \beta|$ and therefore

$$\begin{aligned}
\cos^2(\gamma) &\leq \cos^2(\alpha - \beta) \\
&= (\cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta))^2 \\
&= 2\cos^2(\alpha)\cos^2(\beta) + 1 - \cos^2(\alpha) - \cos^2(\beta) + 2\sin(\alpha)\cos(\alpha)\sin(\beta)\cos(\beta).
\end{aligned}$$

Hence, we get

$$\begin{aligned}
\frac{|\mathcal{P}''(u)[v,w]|^2}{|v|^2|w|^2} &\leq |u|^{-4}\left(1 + 2\sin(\alpha)\cos(\alpha)\sin(\beta)\cos(\beta) - \cos^2(\alpha)\cos^2(\beta)\right) \\
&= |u|^{-4}\left(1 + \sin^2(\alpha)\sin^2(\beta) - \cos^2(\alpha+\beta)\right) \\
&\leq |u|^{-4}\left(1 + \sin^4\left(\frac{\alpha+\beta}{2}\right) - \cos^2(\alpha+\beta)\right) \\
&= |u|^{-4}\left(\frac{4}{3} - 3\left(\cos^2\left(\frac{\alpha+\beta}{2}\right) - \frac{1}{3}\right)^2\right) \\
&\leq \frac{4}{3}|u|^{-4}.
\end{aligned}$$

Taking the square root yields the desired result. Equality holds for $\alpha = \beta = \arccos\left(3^{-1/2}\right)$ and $\gamma = 0$. □

B Identities and estimates on sequences

Assume we are given an expression where a variable occurs k times and we want to compute finite differences of it. The goal is to rewrite it as a telescopic sum with finite differences only taken with respect to one occurence of the variable. The following lemma shows that this can be done.

Lemma B.1. *Let $\mathfrak{u}\colon \mathbb{Z}^k \to X$, where X is a vector space and $\mathfrak{v}\colon \mathbb{Z} \to X$ be the diagonal, i.e. $\mathfrak{v}_i := \mathfrak{u}_{i,\dots,i}$ for all $i \in \mathbb{Z}$. Let ∇_j be the difference operator only applied to the j-th coordinate, i.e.*

$$\nabla_j \mathfrak{u}_{i_1,\dots,i_k} := \mathfrak{u}_{i_1,\dots,i_j+1,\dots,i_k} - \mathfrak{u}_{i_1,\dots,i_j,\dots,i_k}.$$

Then we have for $i \in \mathbb{Z}$

$$(\nabla^l \mathfrak{v})_i = \sum_{i=a_0\le a_1\cdots\le a_k=i+l} \frac{l!}{\prod_{m=1}^k (a_m - a_{m-1})!} \nabla_1^{a_1-a_0} \dots \nabla_k^{a_k-a_{k-1}} \mathfrak{u}_{a_0,\dots,a_{k-1}}.$$

Proof. The right hand side of the expression above is a sum of terms of the form $\mathfrak{u}_{i_1,\dots,i_k}$ with $i = i_0 \le i_1 \le i_2 \cdots \le i_k \le i_{k+1} = i + l$. Note that it is enough to prove that the coefficients are

$$\begin{cases} (-1)^{l-j}\binom{l}{j} & \text{if } i_1 = \cdots = i_k = i + j \\ 0 & \text{otherwise} \end{cases}.$$

We have

$$\begin{aligned} & \sum_{a_1=i_1}^{i_2} \cdots \sum_{a_{k-1}=i_{k-1}}^{i_k} \frac{l!}{\prod_{m=1}^k (a_m - a_{m-1})!} \prod_{m=1}^k (-1)^{a_m - i_m} \frac{(a_m - a_{m-1})!}{(i_m - a_{m-1})!(a_m - i_m)!} \\ = \; & \frac{l!(-1)^{a_k - i_k}}{(a_k - i_k)!(i_1 - a_0)! \prod_{m=1}^{k-1} (i_{m+1} - i_m)!} \prod_{m=1}^{k-1} \sum_{a_m=i_m}^{i_{m+1}} (-1)^{a_m - i_m} \binom{i_{m+1} - i_m}{a_m - i_m} \\ = \; & \frac{l!(-1)^{a_k - i_k}}{\prod_{m=0}^k (i_{m+1} - i_m)!} \prod_{m=1}^{k-1} (1-1)^{i_{m+1} - i_m} \\ = \; & \begin{cases} (-1)^{l-j}\binom{l}{j} & \text{if } i_1 = \cdots = i_k = i + j \\ 0 & \text{otherwise} \end{cases}. \quad \square \end{aligned}$$

In Section 3.5 we work with a sequence $\mathfrak{b}$ with $\sum_{i\in\mathbb{Z}} \mathfrak{b}_i = 1$ and we want to estimate the expression $\|\mathfrak{b} * \mathfrak{u} - \mathfrak{u}\|_{\ell^\infty}$. The following lemma shows that $\mathfrak{b} * \mathfrak{u} - \mathfrak{u}$ can be rewritten in terms of $\nabla\mathfrak{u}$. This will allow us to estimate $\|\mathfrak{b} * \mathfrak{u} - \mathfrak{u}\|_{\ell^\infty}$.

Lemma B.2. *Let $\mathfrak{b} \in \mathbb{R}^{\mathbb{Z}}$ be an invertible sequence with finite support and $\sum_{i\in\mathbb{Z}} \mathfrak{b}_i = 1$. Then there exist a sequence $\mathfrak{a} \in \mathbb{R}^{\mathbb{Z}}$ with finite support such that*

$$\mathfrak{b} * \mathfrak{u} = \mathfrak{u} + \mathfrak{a} * \nabla\mathfrak{u}$$

Furthermore, if $\mathfrak{b}^{-1}$ is an inverse of $\mathfrak{b}$ we have

$$\mathfrak{b}^{-1} * \mathfrak{u} = \mathfrak{u} - \mathfrak{b}^{-1} * \mathfrak{a} * \nabla\mathfrak{u}.$$

Proof. Let the support of $\mathfrak{b}$ be in $\{-S,\dots,S\}$. We define

$$\mathfrak{a}_i := \begin{cases} -\sum_{j=i}^{S} \mathfrak{b}_j & 1 \le i \le S \\ \sum_{j=-S}^{i-1} \mathfrak{b}_j & -S+1 \le i \le 0 \\ 0 & \text{otherwise} \end{cases}$$

and get by the assumption

$$(\nabla \mathfrak{a})_i = \begin{cases} \mathfrak{b}_i & i \neq 0 \\ -\sum_{i\neq 0} \mathfrak{b}_i = \mathfrak{b}_0 - 1 & i = 0 \end{cases}.$$

Note that we have $\nabla(\mathfrak{b} * \mathfrak{u}) = \mathfrak{b} * \nabla\mathfrak{u} = \nabla\mathfrak{b} * \mathfrak{u}$. Hence,

$$\mathfrak{u} + \mathfrak{a} * \nabla\mathfrak{u} = \mathfrak{u} + \nabla\mathfrak{a} * \mathfrak{u} = \mathfrak{u} + \mathfrak{b} * \mathfrak{u} - \mathfrak{u} = \mathfrak{b} * \mathfrak{u}.$$

The second statement follows by multiplying with $\mathfrak{b}^{-1}$ from the left on both sides. □

To simplify the notation we will omit the convolution sign "*" in the following lemma.

Lemma B.3. *Let $\mathfrak{u} \in M^{\mathbb{Z}}$ and $\mathfrak{v} \in \left(\mathbb{R}^K\right)^{\mathbb{Z}}$. We have*

$$\|\mathcal{P}\mathfrak{b}\mathcal{P}(\mathfrak{u}+\mathfrak{v}) - \mathcal{P}\mathfrak{b}(\mathfrak{u}+\mathfrak{v})\|_{\ell^\infty} \lesssim \|\mathfrak{v}\|_{\ell^\infty} \left(\|\mathfrak{v}\|_{\ell^\infty} + \|\nabla\mathfrak{u}\|_{\ell^\infty}\right).$$

Proof. Let $i \in \mathbb{Z}$ and $x := \mathcal{P}(\mathfrak{b}\mathfrak{u})_i$. Using the Taylor expansion of $\mathcal{P}$ at $(\mathfrak{b}\mathfrak{u})_i$ we get

$$\left|\mathcal{P}(\mathfrak{b}(\mathfrak{u}+\mathfrak{v}))_i - (\mathcal{P}(\mathfrak{b}\mathfrak{u})_i + P_{T_xM}(\mathfrak{b}\mathfrak{v})_i)\right| \lesssim \left|(\mathfrak{b}\mathfrak{v})_i\right|^2.$$

where P_{T_xM} denotes the orthogonal projection onto the tangent space T_xM. Hence

$$\|\mathcal{P}\mathfrak{b}(\mathfrak{u}+\mathfrak{v}) - (\mathcal{P}(\mathfrak{b}\mathfrak{u}) + P_{T_{\mathcal{P}\mathfrak{b}\mathfrak{u}}M}\mathfrak{b}\mathfrak{v})\|_{\ell^\infty} \lesssim \|\mathfrak{v}\|_{\ell^\infty}^2.$$

Similarly

$$\|\mathcal{P}\mathfrak{b}\mathcal{P}(\mathfrak{u}+\mathfrak{v}) - (\mathcal{P}(\mathfrak{b}\mathfrak{u}) + P_{T_{\mathcal{P}\mathfrak{b}\mathfrak{u}}M}P_{T_{\mathfrak{u}}M}\mathfrak{b}\mathfrak{v})\|_{\ell^\infty} \lesssim \|\mathfrak{v}\|_{\ell^\infty}^2.$$

Using $P_{T_xM}^2 = P_{T_xM}$ for all $x \in M$ and the triangle inequality we get

$$\begin{aligned} \|\mathcal{P}\mathfrak{b}\mathcal{P}(\mathfrak{u}+\mathfrak{v}) - \mathcal{P}\mathfrak{b}(\mathfrak{u}+\mathfrak{v})\|_{\ell^\infty} &\lesssim \|P_{T_{\mathcal{P}\mathfrak{b}\mathfrak{u}}M}(P_{T_{\mathcal{P}\mathfrak{b}\mathfrak{u}}M} - P_{T_{\mathfrak{u}}M})\mathfrak{b}\mathfrak{v}\|_{\ell^\infty} + \|\mathfrak{v}\|_{\ell^\infty}^2 \\ &\lesssim \|P_{T_{\mathcal{P}\mathfrak{b}\mathfrak{u}}M} - P_{T_{\mathfrak{u}}M}\|_{\ell^\infty} \|\mathfrak{v}\|_{\ell^\infty} + \|\mathfrak{v}\|_{\ell^\infty}^2 \end{aligned}$$

Using the smoothness of $\mathcal{P}$ we get

$$\|P_{T_{\mathcal{P}\mathfrak{b}\mathfrak{u}}M} - P_{T_{\mathfrak{u}}M}\|_{\ell^\infty} \lesssim \|\mathcal{P}\mathfrak{b}\mathfrak{u} - \mathfrak{u}\|_{\ell^\infty}.$$

Using $|\mathcal{P}x - \mathcal{P}y| \lesssim |x-y|$ and Lemma B.2 we get

$$\|\mathcal{P}\mathfrak{b}\mathfrak{u} - \mathfrak{u}\|_{\ell^\infty} \lesssim \|\mathfrak{b}\mathfrak{u} - \mathfrak{u}\|_{\ell^\infty} = \|c\nabla\mathfrak{u}\|_{\ell^\infty} \lesssim \|\nabla\mathfrak{u}\|_{\ell^\infty}.$$

Combining these inequalities we get the desired estimate. □

C Estimates on the discrete harmonic energy

The approximation error of the harmonic energy in dimension 1 is given by the following Lemma.

Lemma C.1. *Let M be a Riemannian submanifold of $\mathbb{R}^n$ and d a metric on M with $d(u,v) = |v-u| + \mathcal{O}(|v-u|^3)$. Then for $u \in C^3([0,h], M)$ we have*

$$\int_0^h |u'(x)|^2 dx - h^{-1} d^2(u(h), u(0)) = \mathcal{O}\left(h^3 \|u\|_{C^3}^2\right).$$

Proof. Note that it is enough to prove the statement for $d(u,v) = |v-u|$. Taylor expansion of u' yields

$$u'(x) = u'(0) + u''(0)x + \mathcal{O}\left(|x|^2 |u|_{C^3}\right).$$

Hence we have

$$\begin{aligned} \int_0^h |u'(x)|^2 dx &= \int_0^h |u'(0)|^2 + 2x\langle u'(0), u''(0)\rangle + \mathcal{O}(|x|^2 \|u\|_{C^3}^2) dx \\ &= |u'(0)|^2 h + h^2 \langle u'(0), u''(0)\rangle + \mathcal{O}(h^3 \|u\|_{C^3}^2). \end{aligned}$$

Taylor expansion of u yields

$$u(x) = u(0) + u'(0)x + \frac{1}{2} u''(0)x^2 + \mathcal{O}\left(|x|^3 |u|_{C^3}\right).$$

Hence, we have

$$\begin{aligned} h^{-1}|u(h) - u(0)| &= h^{-1}|u'(0)h + \frac{1}{2}u''(0)h^2|^2 + \mathcal{O}(h^3 \|u\|_{C^3}^2) \\ &= |u'(0)|^2 h + h^2 \langle u'(0), u''(0)\rangle + \mathcal{O}(h^3 \|u\|_{C^3}^2). \end{aligned}$$

Taking the difference of both approximations yields the desired result. □

To estimate the harmonic energy by the discrete harmonic energy of sphere-valued functions we will need the following lemma.

Lemma C.2. *Let $V = \{0,1\}^s$ be the s-dimensional hypercube, $E \subset V \times V$ the edges of the hypercube, $x \in [0,1]^s$ and for $i \in \{0,1\}^s$ the function ϕ_i as defined in Section 4.2.2 and $a = (a_i)_{i \in V} \in (S^n)^V$. Then we have*

$$\left| \sum_{i \in V} \phi_i(x) a_i \right|^2 \geq 1 - \frac{1}{4} \sum_{(i,j) \in E} \phi_{i,j}(x) |a_i - a_j|^2,$$

where $\phi_{i,j} := \phi_i + \phi_j$.

Proof. We prove the statement by induction on s. The case $s = 1$ follows from

$$|ta_1+(1-t)a_2|^2 = t^2|a_1|^2+(1-t)^2|a_2|^2-2t(1-t)\langle a_1, a_2\rangle = 1-t(1-t)|a_1-a_2|^2 \geq 1-\frac{1}{4}|a_1-a_2|^2,$$

where we used that $t(1-t) \leq \frac{1}{4}$. To prove the induction step $s \to s+1$ we show that for all $t \in [0,1]$ and $a, b \in V$ we have

$$\left|\sum_{i\in V} \phi_i(x)(ta_i + (1-t)b_i)\right|^2 \geq 1-\frac{1}{4}\sum_{(i,j)\in E} \phi_{i,j}(x)(t|a_i-a_j|^2+(1-t)|b_i-b_j|^2)-\frac{1}{4}\sum_{i\in V}\phi_i(x)|a_i-b_i|^2.$$

By the triangle and the Cauchy–Schwarz inequality we have with $u := \sum_{i\in V} \phi_i(x)a_i$ and $v := \sum_{i\in V} \phi_i(x)b_i$ that

$$|v-u|^2 = \left|\sum_{i\in V}\phi_i(x)(b_i-a_i)\right|^2 \leq \left(\sum_{i\in V}\phi_i(x)\right)\left(\sum_{i\in V}\phi_i(x)|b_i-a_i|^2\right) = \sum_{i\in V}\phi_i(x)|b_i-a_i|^2.$$

Hence we have using the induction hypothesis and $t(1-t) \leq \frac{1}{4}$ that

$$\begin{aligned}
&\left|\sum_{i\in V} \phi_i(x)(ta_i + (1-t)b_i)\right|^2 \\
= \;& t^2|u|^2 + (1-t)^2|v|^2 + 2t(1-t)\langle u, v\rangle \\
= \;& t|u|^2 + (1-t)|v|^2 - t(1-t)|u-v|^2 \\
\geq \;& 1-\frac{1}{4}\sum_{(i,j)\in E}\phi_{i,j}(x)(t|a_i-a_j|^2+(1-t)|b_i-b_j|^2)-\frac{1}{4}\sum_{i\in V}\phi_i(x)|a_i-b_i|^2. \quad \square
\end{aligned}$$

Publications

- P. Grohs and M. Sprecher, Total variation regularization on Riemannian manifolds by iteratively reweighted minimization, 2016, Information and Inference
- P. Grohs and M. Sprecher and Thomas Yu, Scattered Manifold-Valued Data Approximation, 2016, Numerische Mathematik

Bibliography

[1] T. J. Abatzoglou. The minimum norm projection on C^2-manifolds in $\mathbb{R}^n$. *Transaction of the American Mathematical Society*, 243:115–122, 1978.

[2] P.-A. Absil. Projection-like retractions on matrix manifolds. *SIAM Journal on Optimization*, 22:135–185, 2012.

[3] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

[4] P.-A. Absil, R. Mahony, and J. Trumpf. An extrinsic look at the Riemannian Hessian. In *Geometric Science of Information*, volume 8085 of *Lecture Notes in Computer Science*, pages 361–368. Springer, 2013.

[5] F. Alouges. A new algorithm for computing liquid crystal stable configurations: the harmonic mapping case. *SIAM Journal on Numerical Analysis*, 34(5):1708–1726, 1997.

[6] T. Aykin and A. J. B. Babu. Multifacility location problems on a sphere. *International Journal of Mathematics and Mathematical Science*, 10(3):583–596, 1987.

[7] W. Ballmann. *Lectures on spaces of nonpositive curvature*, volume 25 of *DMV Seminar*. Birkhäuser Verlag, Basel, 1995.

[8] A. Barmpoutis, B. Jian, and B. C. Vemuri. Adaptive kernels for multi-fiber reconstruction. *In LNCS 5636 (Springer) Proceedings of IPMI09: Information Processing in Medical Imaging*, pages 338–349, 2009.

[9] D. Braess. *Finite Elemente-Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie*. Springer, 2007.

[10] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*, volume 1. Springer, 2011.

[11] M. R. Bridson and A. Häfliger. *Metric spaces of non-positive curvature*. Springer, 1999.

[12] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[13] L. Carlitz. Eulerian numbers and polynomials of higher order. *Duke Mathematical Journal*, 27(3):401–423, 1960.

[14] É. Cartan. *La géométrie des espaces de Riemann*. Gauthier-Villars, 1925.

[15] M. Cavegn. Total variation regularization for geometric data. *Master Thesis, ETH Zürich*, 2013.

[16] O. Christiansen, T.-M. Lee, J. Lie, U. Sinha, and T. F. Chan. Total variation regularization of matrix-valued images. *Internation Journal of Biomedical Imaging*, 2007.

[17] O. Coulon, D. C. Alexander, and S. R. Arridge. Diffusion tensor magnetic resonance image regularization. *Medical Image Analysis*, 8(1):47–67, 2004.

[18] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 63:1–38, 2010.

[19] P. Debus. A fast C++ template library for total variation minimization of manifold-valued two- and three-dimensional images. `https://github.com/pdebus/MTVMTL`, 2015.

[20] A. DeSimone, R.V. Kohn, S. Müller, and F. Otto. Recent analytical developments in micromagnetics. *The Science of Hysteresis. Vol. 2 Physical Modeling, Micromagnetics, and Magnetization Dynamics*, 2, 2006.

[21] U. R. Dhar and J. R. Rao. Domain approximation method for solving multifacility location problems on a sphere. *The Journal of the Operational Research Society*, 33(7):pp. 639–645, 1982.

[22] G. Frobenius. Über die bernoullischen Zahlen und die Eulerschen Polynome. *Sitzungsberichte der Preussischen Akademie der Wissenschaften / Physikalisch-Mathematische Klasse*, pages 809–847, 1910.

[23] P. Grohs. Finite elements of arbitrary order and quasiinterpolation for Riemannian data. *IMA Journal of Numerical Analysis*, 33(3):849–874, 2013.

[24] P. Grohs, H. Hardering, and O. Sander. Optimal a priori discretization error bounds for geodesic finite elements. *Foundations of Computational Mathematics*, 15:1357–1411, 2014.

[25] N. J. Higham. Computing the polar decomposition with applications. *SIAM Journal on Scientific and Statistical Computing*, 7:1160–1174, 1986.

[26] W. Jäger and H. Kaul. Uniqueness and stability of harmonic maps and their Jacobi fields. *Manuscripta Mathematica*, 28:269–291, 1979.

[27] A. Jonsson and H. Wallin. A Whitney extension theorem in L^p and Besov spaces. *Ann. Inst. Fourier*, pages 139–192, 1978.

[28] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, 1977.

[29] D. Le Bihan, J.-F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriat. Diffusion tensor imaging: Concepts and applications. *Magnetic Resonance Imaging*, 13(4):534–546, 2001.

[30] J. Lellmann, E. Strekalovskiy, S. Koetter, and D. Cremers. Total variation regularization for functions with values in a manifold. *IEEE International Conference on Computer Vision*, pages 2944–2951, 2013.

[31] R. Linker, O. Cohen, and A. Naor. Determination of the number of green apples in RGB images recorded in orchards. *Computers and Electronics in Agriculture*, 81:45–57, 2012.

[32] M. Marcus and V. J. Mizel. Every superposition operator mapping one Sobolev space into another is continuous. *Journal of Functional Analysis*, 33:217–229, 1979.

[33] S. Mazur and S. Ulam. Sur les transformations isométriques d'espaces vectoriels normés. *C. R. Acad. Sci. Paris*, 194:946–948, 1932.

[34] J. Milnor. *Topology from the differentiable viewpoint.* Princeton University Press, 1976.

[35] M. Moakher and M. Zéraï. The Riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data. *Journal of Mathematical Imaging and Vision*, 40(2):171–187, 2011.

[36] W. Müller. *Numerische Analyse und Parallele Simulation von nichtlinearen Cosserat-Modellen.* PhD thesis, Karlsruher Institut für Technologie, 2009.

[37] I. Münch. *Ein geometrisch und materiell nichtlineares Cosserat-Model — Theorie, Numerik und Anwendungsmöglichkeiten.* PhD thesis, Universität Karlsruhe, 2007.

[38] M. Nägelin. Total variation regularization for tensor valued images. *Bachelor Thesis, ETH Zürich*, 2014.

[39] J. Nash. The imbedding problem for Riemannian manifolds. *Annals of Mathematics*, 63(1):20–63, 1956.

[40] P. Neff. A geometrically exact Cosserat shell-model including size effects, avoiding degeneracy in the thin shell limit. Existence of minimizers for zero Cosserat couple modulus. *Mathematical Models and Methods in Applied Sciences*, 17(3):363–392, 2007.

[41] L. Nirenberg. On elliptic partial differential equations. *Annali della Scuola Normale Superiore di Pisa*, 3:115–162, 1959.

[42] J. Nitsche. Ein Kriterium für die Quasi-optimalität des Ritzschen Verfahrens. *Numerische Mathematik*, 11:346–348, 1968.

[43] M. J. D. Powell. On search directions for minimization algorithms. *Mathematical Programming*, 4(1):193–201, 1973.

[44] P. Rodríguez and B. Wohlberg. An iteratively reweighted norm algorithm for minimization of total variation functionals. *IEEE Signal Processing Letters*, 14(12):948 – 951, 2007.

[45] T. Runst and W. Sickel. *Sobolev spaces of fractional order, Nemytskij operators, and nonlinear partial differential equations*, volume 3. de Gruyter, 1996.

[46] O. Sander. Geodesic finite elements for Cosserat rods. *International Journal for Numerical Methods in Engineering*, 82(13):1645–1670, 2010.

[47] G. Steidl, S. Setzer, B. Popilka, and B. Burgeth. Restoration of matrix fields by second-order cone programming. *Computing*, 81(2-3):161–178, 2007.

[48] G. Strang and G. Fix. *An analysis of the finite element method*. Prentice Hall, 1973.

[49] K.T. Sturm. Probability measures on metric spaces of nonpositive curvature. *Contemporary Mathematics*, 338:357–390, 2003.

[50] P. E. Trahanias, D. Karakos, and A. N. Venetsanopoulos. Directional processing of color images: theory and experimental results. *IEEE Transactions on Image Processing*, 5(6):868–880, 1996.

[51] D. Tschumperlé and R. Deriche. Diffusion tensor regularization with constraints preservation. *IEEE Computer Society Press*, 1:948–953, 2001.

[52] A. Tsuyoshi. Log majorization and complementary Golden-Thompson type inequalities. *Linear Algebra and its Applications*, 197-198:113–131, 1994.

[53] A. Weinmann, L. Demaret, and M. Storath. Total variation regularization for manifold-valued data. *SIAM Journal on Imaging Sciences*, 7(4):2226–2257, 2014.

[54] M. Wellershof. On the convergence rate of gradient descent for the computation of Riemannian averages. *Bachelor Thesis, ETH Zürich*, 2015.

[55] N. Wiener. Tauberian theorem. *Annals of Mathematics*, 33(1):1–100, 1932.

Curriculum Vitae

Personal details

Name	Markus Sprecher
Date of birth	August 30, 1986
Place of birth	Grabs, Switzerland
Citizenship	Swiss

Education

08/2012–01/2016	PhD studies in Mathematics at ETH Zürich, Switzerland
04/2011-2012	Practica at Mettler Toledo, Nänikon ZH
10/2006–04/2011	Studies in Mathematics at ETH Zürich
05/2005-05/2006	Military Service
07/2001–07/2005	Secondary school at Gymnasium Sargans, Switzerland