Nonparametric Regression between General Riemannian Manifolds[*]

Florian Steinke[†], Matthias Hein[‡], and Bernhard Schölkopf[§]

Abstract. We study nonparametric regression between Riemannian manifolds based on regularized empirical risk minimization. Regularization functionals for mappings between manifolds should respect the geometry of input and output manifold and be independent of the chosen parametrization of the manifolds. We define and analyze the three most simple regularization functionals with these properties and present a rather general scheme for solving the resulting optimization problem. As application examples we discuss interpolation on the sphere, fingerprint processing, and correspondence computations between three-dimensional surfaces. We conclude with characterizing interesting and sometimes counterintuitive implications and new open problems that are specific to learning between Riemannian manifolds and are not encountered in multivariate regression in Euclidean space.

Key words. harmonic map, biharmonic map, Eells energy, regularized empirical risk minimization, thin-plate spline

AMS subject classifications. 41A15, 49Q99, 68T10

DOI. 10.1137/080744189

1. Introduction. The concept of a Riemannian manifold is a useful abstraction for many data types in image processing and beyond. For example, the human perception of color values can be modelled via a color circle [38]. Similarly, circular structure is found for interpherometric measurements in SAR (synthetic aperture radar) images [25]. In image and video sequence formation, classic manifolds such as rotations and combinations thereof are ubiquitous, and more complex Riemannian manifolds appear as surfaces of three-dimensional objects [40]. More abstractly, it is often useful in image processing to study whole sets of images themselves as Riemannian manifolds [42], or similarly in shape processing, to define a Riemannian structure on the set of all shapes [19, 9].

In this paper we focus on a classic statistical problem with such data, namely regression between two Riemannian manifolds. Nonparametric regression subsumes interpolation, extrapolation, and smoothing as special cases, which are ubiquitous problems in image processing and image formation. The common problem is to learn from given pairs of input and output examples a mapping between the input and the output space.

The regression problem where input and output domains are Riemannian manifolds is quite distinct from standard multivariate regression between Euclidean spaces. It produces a large variety of new and interesting features, some of which we will describe in this paper. One

[†]Max Planck Institute for Biological Cybernetics, 72706 Tübingen, Germany, and Siemens Corporate Technology, 81739 München, Germany (Florian.Steinke@siemens.com).
[‡]Saarland University, 66041 Saarbrücken, Germany (hein@cs.uni-sb.de).
[§]Max Planck Institute for Biological Cybernetics, 72706 Tübingen, Germany (bs@tuebingen.mpg.de).

fundamental problem of using traditional regression methods for manifold-valued regression is that most standard regression schemes assume that the output space is linear. It thus makes sense to linearly combine simple basis functions, since the addition of function values is still an element of the target space. While this approach still works for manifold-valued input, it is no longer feasible if the output space is a manifold, as general Riemannian manifolds do not have linear structure.

One way that one can still learn manifold-valued mappings using standard regression techniques is to learn mappings directly into charts of the manifold. Another approach is to use an embedding of the manifold in Euclidean space and utilize back-projections onto the manifold. While both approaches yield manifold-valued mappings, the solution will depend on the chart or embedding, respectively, and in particular will not respect the geometric local relationships of the manifold, since close points in Euclidean space need not be close in the geometry of the manifold.

Our approach for regression between manifolds is based on regularized empirical risk minimization, directly influencing the smoothness of the learned mappings via the regularizer. We describe the construction of general regularization functionals based on the derivatives of mappings between Riemannian manifolds and discuss in more detail three specific functionals, namely the harmonic, biharmonic, and novel Eells energy, which can be seen as a generalization of the thin-plate-spline energy. One important property of a regularization functional is its null space, the set of mappings which are not penalized. Interestingly, in the case of the Eells energy the null space turns out to be the set of totally geodesic maps. These maps can be seen as a proper generalization of the set of linear mappings to the case of Riemannian manifolds. Computing the regularization functionals straightforwardly from their definition is often quite difficult. However, if the input and/or output manifold can be embedded isometrically in Euclidean space, the regularization functionals can be rewritten in a much simpler form. From this formulation we construct a relatively simple yet very versatile implementation, which allows us to demonstrate regression between manifolds in four applications. First, we show the differences among the three regularizers for two interpolation tasks on the sphere. We continue with an application to fingerprint reconstruction and finally apply the presented framework in a realistic surface registration problem. We conclude the article by discussing some challenging yet very interesting new mathematical and statistical questions which arise due to the non-Euclidean structure of input and/or output space.

The general learning setup is described in section 2. In section 3 we define regularization functionals for manifold-valued mappings, followed by a discussion of their properties in section 4. In section 5 we provide expressions of the regularization functionals in terms of embeddings, which turns out to be crucial for efficient implementation, our version of which is described in section 6. Experimental results are shown in section 7, and we conclude with some interesting aspects and open problems arising for learning between Riemannian manifolds in section 8.

1.1. Related work. Riemannian manifolds are commonly used in so-called manifold learning, where either only the input domain is considered to be a manifold [3] or a description of the manifold itself is learned [42]. In both cases the manifold is unknown, and only a sample of points from this manifold is given. Instead, the focus in this work is to learn a predictor

from given pairs of input/output examples lying on *known* input and output manifolds.

For regression with manifold-valued output there are classic methods for spherical data [13]. Recently, k-nearest neighbor [17, 6], Nadaraya–Watson [9], and wavelet [35] types estimators have been adapted for this task. In contrast, our work is based on differential energies for mappings between general Riemannian manifolds. It unifies and extends previous approaches in that direction in various ways. The harmonic [12, 44] and biharmonic [27] energies have been studied extensively in the differential geometry community, but less so in a learning context. Close to our setting are [14, 30, 22, 7]. All of these consider the problem of learning a curve in the output manifold; that is, in contrast to our approach, the input domain is constrained to be one-dimensional and Euclidean. Interpolation is performed in [14, 30] with a regularizer that penalizes second-order derivatives, whereas [7] proposes regularization functionals of arbitrary order. Approximation is analyzed in [22], but only a first-order regularizer is used. All these approaches fix start- and endpoints of the curve. The closest in spirit to our approach is [26], where the harmonic energy is used in an approximation setting.

Parts of the present work have been previously presented at conferences [40, 39]. This paper summarizes our previous work in a unified and self-contained way. Additionally, we here compare our approach with other potential regularization energies such as the harmonic and the biharmonic energy and provide arguments as to why the proposed intrinsic approach is superior to an extrinsic treatment [16]. Moreover, we present interesting and challenging problems that arise in nonparametric regression between manifolds and that are not present in the Euclidean case. Finally, in the appendix we derive the Euler–Lagrange equations for all discussed energies.

1.2. Notation. Throughout the article we use the following notation. M is always the input manifold, N the target manifold, and $\phi : M \to N$ is a mapping from input to target manifold. The dimensions of M and N are m and n, and x and y are coordinates in M and N. Moreover, we will use the Einstein summation convention and Penrose's abstract index notation; see [46, Chapter 2.4]. "Abstract" indices indicate only the tensor type; they should not be confused with the indices for the components. For example, a two-times covariant tensor h is written as h_{ab}, and the coordinate representation would be $h_{ab} = h_{\mu\nu}\, dx_a^\mu \otimes dx_b^\nu$. In general, we use Greek letters for components (α, β, γ for components in M and μ, ν, ρ for components in N) and Latin for abstract indices (a, b, c for indices in M and r, s, t in N). We denote by g_{ab}, h_{ab} the metrics on M and N, and by ${}^M\nabla$ and ${}^N\nabla$ the Levi–Civita connections on M and N with corresponding Christoffel symbols ${}^M\Gamma^\alpha_{\beta\gamma}$ and ${}^N\Gamma^\rho_{\nu\mu}$. We follow [21] in defining the Riemannian curvature tensor $R : \otimes^3 TM \otimes T^*M \to \mathbb{R}$ as $\nabla_a\nabla_b Z^c - \nabla_b\nabla_a Z^c = R_{abd}{}^c\, Z^d$. As usual, $\otimes$ denotes the tensor product.

2. Regularized empirical risk minimization for manifold-valued regression. Given a set of K training pairs (X_i, Y_i) with $X_i \in M$ and $Y_i \in N$, we would like to learn a mapping $\phi : M \to N$. This learning problem reduces to standard multivariate regression if M and N are both Euclidean spaces $\mathbb{R}^m$ and $\mathbb{R}^n$, and to regression on a manifold if at least N is Euclidean. We propose to use regularized empirical risk minimization, which can be formulated in our setting as

$$(2.1) \qquad \underset{\phi \in C^\infty(M,N)}{\arg\min}\; \frac{1}{K}\sum_{i=1}^{K} L\big(Y_i, \phi(X_i)\big) + \lambda\, S(\phi),$$

where $C^\infty(M, N)$ denotes the set of smooth mappings ϕ between M and N, $L : N \times N \to \mathbb{R}_+$ is the loss function, $\lambda \in \mathbb{R}_+$ the regularization parameter, and $S : C^\infty(M, N) \to \mathbb{R}$ the regularization functional. The regularization functional should measure the complexity of the mapping ϕ; the proper definition of such a functional will be the topic of the next section.

2.1. Loss function. In multivariate regression, $f : \mathbb{R}^m \to \mathbb{R}^n$, the most common loss function is the squared Euclidean distance of $f(X_i)$ and Y_i, $L(Y_i, f(X_i)) = \|Y_i - f(X_i)\|^2_{\mathbb{R}^n}$. A direct generalization to a loss function on a Riemannian manifold N is to use the squared geodesic distance in N, $L(Y_i, \phi(X_i)) = d_N^2(Y_i, \phi(X_i))$. The correspondence to the multivariate case can be seen from the fact that $d_N(Y_i, \phi(X_i))$ is the length of the shortest path between Y_i and $\phi(X_i)$ in N, as the norm $\|f(X_i) - Y_i\|$ is the length of the shortest path, namely the length of the straight line, between $f(X_i)$ and Y_i in $\mathbb{R}^n$. Naturally, taking the pth power of the geodesic distance as well as any other function $\Theta : \mathbb{R}_+ \to \mathbb{R}_+$ of the geodesic distance is also possible.

In general, we assume our problem to be in a statistical setting (however, the framework works also if this is not the case), where the given input/output pairs (X_i, Y_i) are i.i.d. (independent and identically distributed) samples from a probability measure P on $\mathcal{X} \times \mathcal{Y}$. The setting we have in mind is that our data is perturbed by noise in the output space. In multivariate regression it is well known that, using the squared Euclidean distance as a loss function, $L(y_i, f(x_i)) = \|y_i - f(x_i)\|^2_{\mathbb{R}^n}$, the Bayes optimal predictor f^*, that is, the function f^* minimizing the expected loss,

$$f^* = \underset{f \text{ measurable}}{\arg\min} \ \mathbb{E} \|Y - f(X)\|^2 = \underset{f \text{ measurable}}{\arg\min} \ \mathbb{E}_X \mathbb{E}_{Y|X}[\|Y - f(x)\|^2 \mid X],$$

is given by the regression function $f^*(x) = \mathbb{E}[Y|X = x]$. The regression function $f^*(x)$ is uniquely determined (almost everywhere) since the risk functional is strictly convex in $f(x)$.

Naturally, the question arises of which is the Bayes optimal mapping $\phi^* : M \to N$ in the manifold case; that is, using the squared geodesic distance in N as a loss measure, which map ϕ^* minimizes the expected loss,

$$\phi^* = \underset{\phi:M\to N, \ \phi \text{ measurable}}{\arg\min} \ \mathbb{E}_X \mathbb{E}_{Y|X}[d_N^2(Y, \phi(X)) \mid X],$$

where we assume that $\mathbb{E} \, d_N^2(Y, \phi(X)) < \infty$ for some measurable $\phi : M \to N$. Using the factorization of the probability measure P, the Bayes optimal mapping ϕ^* can be determined pointwise,

$$\phi^*(x) = \underset{p \in N}{\arg\min} \, \mathbb{E}[d_N^2(Y, p) \mid X = x] = \underset{p \in N}{\arg\min} \, \underbrace{\int_N d^2(y, p) \, d\mu_x(y)}_{\equiv F(p)},$$

where $d\mu_x$ is the conditional probability measure for Y given $X = x$. The global minimizer of $F(p)$ is called the Frechét or Karcher mean [17].[1] It is the direct generalization of the mean

---

[1] In some cases the set of all local minimizers is denoted as the Frechét mean set, and the mean is called unique if there exists only one global minimizer.

in Euclidean space to general metric spaces. Unfortunately, it need no longer be unique as in the Euclidean case. A simple example is the sphere with a uniform probability measure on it. In this case every point p on the sphere attains the same value $F(p)$, and thus the global minimum is nonunique. We refer to [18, 4] for more information on the conditions under which uniqueness can be proven.

2.2. Function space. Note, that in (2.1) we constrain ϕ to be smooth. However, it is well known for Euclidean output spaces that optimizing a regularization functional based on derivatives over the set of differentiable functions may in some situations yield trivial non-continuous solutions; see Sobolev's embedding theorem [2]. However, one can give conditions on the dimension of the input space or alternatively on the order of the penalized derivatives for which such trivial solutions are not possible. While the generalization of this theory to the manifold-valued case is beyond the scope of this paper, we discuss these issues and the theory of spaces of manifold-valued mappings in more detail in section 8.1. As Euclidean output is a special case of manifold-valued output, it is likely that the restrictions encountered there translate to the more general problem discussed in this paper. In all the experiments of this work, these restrictions are satisfied. For simplicity, we assume in the following that all considered mappings are smooth.

3. Regularization functionals for mappings between Riemannian manifolds. We would like to define regularization functionals,

$$S : C^\infty(M, N) \to \mathbb{R}_+,$$

for mappings between two Riemannian manifolds M and N measuring the smoothness of mapping $\phi : M \to N$ via its derivatives. Two objectives should hold for the regularization functional:

1. independence of the parametrization of M and N,
2. intrinsic formulation; that is, the energy should depend only on the geometry of M and N.

The first objective means that the energy should not depend on the coordinate representation of the manifold; e.g., the energy of curves on the sphere should be the same if we represent the sphere in spherical or stereographic coordinates. This can be achieved by formulating the energy in the covariant language of differential geometry. The second requirement is that the energy should depend only on the geometry of M and N; that is, only *intrinsic* properties of M and N should matter. In particular, if M and N are isometrically embedded in Euclidean space like the sphere S^2 in $\mathbb{R}^3$ or SO_3 in $\mathbb{R}^{3\times3}$, no properties of the ambient spaces should be taken into account, since the embedding is not unique. We will show in section 5.3 that the penalization of components in the ambient space (extrinsic quantities) leads to a notion of smoothness for manifold-valued mappings which contradicts our intuitive expectations.

The remainder of this section requires some technical notions from differential geometry, particularly that of a pull-back connection. For the sake of a clear presentation we have moved the exact definition of this term to Appendix A. The basic properties can be understood also without this knowledge.

Before we discuss general regularization functionals penalizing derivatives of arbitrary order let us begin with the most simple energy functional for manifold-valued mappings. The

differential or Jacobian $d\phi_a^r : T_x M \to T_{\phi(x)} N$ of a mapping $\phi : M \to N$ evaluated at x is given as

$$(3.1) \qquad d\phi_a^r(x) = \frac{\partial \phi^\mu}{\partial x^\alpha} dx_a^\alpha \Big|_x \otimes \frac{\partial^r}{\partial y^\mu} \Big|_{\phi(x)}.$$

It measures the change of the output $\phi(x) \in N$ as one varies x in the input manifold M. This 1-1-tensor can be used to define the most simple differential energy, the so-called harmonic energy.

Definition 3.1. *The* harmonic energy $S_{\text{harmonic}}(\phi)$ *of a mapping* $\phi : M \to N$ *is defined as*

$$(3.2) \qquad S_{\text{harmonic}}(\phi) = \int_M \|d\phi\|_{T_x^* M \otimes T_{\phi(x)} N}^2 \, dV(x)$$
$$= \int_M g^{\alpha\beta}(x) h_{\mu\nu}(\phi(x)) \frac{\partial \phi^\mu}{\partial x^\alpha} \frac{\partial \phi^\nu}{\partial x^\beta} dV(x),$$

where $dV = \sqrt{\det g}\, dx$ *is the volume element of* M.

For standard regression, that is, $M = \mathbb{R}^m$ and $N = \mathbb{R}$, the harmonic energy reduces to

$$S_{\text{harmonic}}(\phi) = \int_{\mathbb{R}^m} \|\nabla \phi\|^2 \, dx.$$

For $m = 1$ this functional in turn reduces to the energy functional of linear splines, and using this energy in approximation or interpolation as in objective (2.1) leads to piecewise linear solutions which are nondifferentiable at the mapped data points $\phi(X_i)$. Similar behavior can be observed for curves on manifolds, that is, for $M = [a, b]$ and N a Riemannian manifold, where

$$S_{\text{harmonic}}(\phi) = \int_a^b \left\| \dot{\phi} \right\|^2 dt$$

with $\dot{\phi}(t) = \frac{d\phi}{dt}(t)$. In this case, minimizers of (2.1) are piecewise geodesic [22].

Since we are generally interested in solutions which have higher smoothness, we have to use higher-order derivatives in the regularizer. In the Euclidean case this is typically done, e.g., by using the thin-plate spline energy $\int_{\mathbb{R}^m} \|Hf\|_F^2 \, dx$, where Hf is the Hessian of $f : \mathbb{R}^m \to \mathbb{R}$ and $\|.\|_F$ the Frobenius norm. Another alternative is the biharmonic regularizer, $\int_{\mathbb{R}^m} |\Delta f|^2 dx$, where $\Delta f = \text{trace}(Hf)$.

For the generalization of regularizers of this type to the case of mappings between manifolds we have to define the second derivative of mappings between Riemannian manifolds, that is, the covariant derivative of the differential $d\phi_a^r$. The problem is here that $d\phi$ "lives" in the cotangent and tangent space, $T_x^* M$ and $T_{\phi(x)} N$, of two different manifolds. Thus we cannot simply use the connection $^M\nabla$ of M. The solution is to use the pull-back connection ∇' defined in Appendix A, which yields a notion of the derivative of a vector field on N with respect to a variation in M, where M and N are connected via $\phi : M \to N$. We then use the pull-back connection for derivatives of vector fields in the target manifold N plus the connection on M for derivatives on the input manifold together in a so-called tensor product connection; see also Appendix A. The pth-order covariant derivative of the differential $d\phi$ will yield the tensor field

$$\nabla'_{a_1} \cdots \nabla'_{a_p} d\phi_{a_{p+1}}^r \ \in \ \otimes^{p+1} T^* M \otimes \phi^{-1} TN,$$

where $\phi^{-1}TN$ is the pull-back bundle; see Definition A.2. This derivative is by definition invariant with respect to parametrization and respects the intrinsic geometry of M and N. For a function $\phi : \mathbb{R}^m \to \mathbb{R}^n$ the pth-order covariant derivative equals

$$\frac{\partial^{p+1}\phi^\mu}{\partial x^{\alpha_1}\cdots\partial x^{\alpha_{p+1}}} dx_{a_1}^{\alpha_1} \otimes \cdots \otimes dx_{a_{p+1}}^{\alpha_{p+1}} \otimes \frac{\partial^r}{\partial y^\mu}.$$

In this form the Euclidean $(p + 1)$-order derivative is covariant, that is invariant under coordinate changes.

We can now define higher-order differential energies. In order to obtain a real-valued regularization functional, we have to define an operation $\Theta : \otimes^{p+1}T^*M \otimes \phi^{-1}TN \to \mathbb{R}_+$. The function Θ usually consists of two steps. One first takes traces in some entries and then the norm or some power of the norm of the resulting tensor. This yields the general regularization functional, $S : C^\infty(M, N) \to \mathbb{R}_+$, defined as

$$(3.3) \qquad S(\phi) = \int_M \Theta\left(\nabla'_{a_1} \cdots \nabla'_{a_p} d\phi^r_{a_p+1}\right) dV.$$

We will illustrate this for second-order differential energies ($p = 1$). The tensor field $\nabla'_b d\phi^r_a$ is given in coordinates (see Appendix A) as

$$(3.4) \qquad \nabla'_b d\phi^r_a = \left[\frac{\partial^2\phi^\mu}{\partial x^\beta \partial x^\alpha} - \frac{\partial\phi^\mu}{\partial x^\gamma}{}^M\Gamma^\gamma_{\beta\alpha} + \frac{\partial\phi^\rho}{\partial x^\alpha}\frac{\partial\phi^\nu}{\partial x^\beta}{}^N\Gamma^\mu_{\nu\rho}\right] dx^\beta_b \otimes dx^\alpha_a \otimes \frac{\partial^r}{\partial y^\mu}.$$

Note that nonvanishing Christoffel symbols of M keep the expression linear in ϕ, whereas nonzero Christoffel symbols of N render the second-order differential a nonlinear operator. This illustrates again why manifold-valued input is easier to handle than manifold-valued output.

For the tensor field $\nabla'_b d\phi^r_a$ we can first take the trace in b and a and then use the squared norm in $T_{\phi(x)}N$, which yields the biharmonic energy.

Definition 3.2. *The* biharmonic energy $S_{\mathrm{biharmonic}}(\phi)$ *is defined as*

$$(3.5) \qquad S_{\mathrm{biharmonic}}(\phi) = \int_M \left\|g^{ba}\nabla'_b d\phi^r_a\right\|^2_{T_{\phi(x)}N} dV(x)$$
$$= \int_M g^{ba}\, g^{cd}\, h_{rs}\, \nabla'_b d\phi^r_a \, \nabla'_c d\phi^s_d \, dV(x).$$

Another possibility is to directly use the squared norm in $T^*_x M \otimes T^*_x M \otimes T_{\phi(x)}N$.

Definition 3.3. *The* Eells energy $S_{\mathrm{Eells}}(\phi)$ *is defined as*

$$S_{\mathrm{Eells}}(\phi) = \int_M \left\|\nabla'_b d\phi^r_a\right\|^2_{T^*_x M \otimes T^*_x M \otimes T_{\phi(x)}N} dV(x)$$
$$(3.6) \qquad = \int_M g^{ac}\, g^{bd}\, h_{rs}\, \nabla'_b d\phi^r_a \, \nabla'_d d\phi^s_c \, dV(x).$$

While the biharmonic energy has been discussed in the differential geometry community (see [27]), the Eells energy has to our knowledge not been studied in differential geometry or

elsewhere before. We have named it after James Eells, recently deceased, who pioneered the study of harmonic maps between Riemannian manifolds [12].

The Eells energy reduces to the thin-plate spline energy in the Euclidean case. If M and N are Euclidean, we obtain

$$S_{\text{Eells}}(\phi) = \int_M g^{\alpha\beta} g^{\gamma\delta} h_{\mu\nu} \frac{\partial^2 \Phi^\mu}{\partial x^\alpha \partial x^\gamma} \frac{\partial^2 \Phi^\nu}{\partial x^\beta \partial x^\delta} \, dV(x),$$

where g and h are the Riemannian metrics corresponding to Euclidean space. This is the parametrization-independent form of the thin-plate spline energy. In Cartesian coordinates we have $g^{\alpha\beta} = \delta^{\alpha\beta}$ and $h_{\mu\nu} = \delta_{\mu\nu}$, where δ is the Kronecker symbol. The Eells energy thus reduces to the standard form of the thin-plate spline energy:

$$(3.7) \qquad S_{\text{Eells}}(\phi) = \sum_{\mu=1}^{n} \int_M \sum_{\alpha,\gamma=1}^{m} \left(\frac{\partial^2 \Phi^\mu}{\partial x^\alpha \partial x^\gamma} \right)^2 dx.$$

For curves ϕ in a manifold N, that is, $M = [a, b]$, the Eells energy and the biharmonic energy are identical,

$$(3.8) \qquad S_{\text{Eells}}(\phi) = S_{\text{biharm.}}(\phi) = \int_a^b \left\| \nabla_{\dot{\phi}(t)} \dot{\phi}(t) \right\|_{T_{\phi(t)}N}^2 dt,$$

where $\dot{\phi}(t) = \frac{\partial}{\partial t} \phi(t)$. Using this energy, we recover the interpolation problem of cubic splines on curved spaces proposed by [14, 30] in our framework (2.1) for $\lambda \to 0$.

4. Properties of the regularization functionals. In this section we describe and compare general properties of the harmonic, biharmonic, and Eells energy and their use as regularizers for regression between two general Riemannian manifolds. We start by describing the null-space of the different functionals, which characterizes the mappings which are *not* penalized; continue with an analysis of the difference between biharmonic and Eells energy; and end with a discussion of why second-order energies are useful in modelling physical systems.

4.1. The null space. The null space of a regularization functional $S(\phi)$, that is, the set $\{\phi \,|\, S(\phi) = 0\}$, is interesting for two reasons. First, the null space consists of the mappings which are not penalized and therefore defines a set of mappings with which we are free to fit the data. In standard regression these are usually linear mappings or polynomials of small degree. The other reason is that, as the regularization parameter λ tends to infinity, the regularized empirical risk minimization problem in (2.1) reduces to

$$(4.1) \qquad \underset{\phi \in C^\infty(M,N)}{\arg\min} \frac{1}{K} \sum_{i=1}^{K} L(Y_i, \phi(X_i)) \quad \text{s.t.} \ S(\phi) = 0.$$

Thus, in this limit the only feasible set of mappings is the null space of S.

Harmonic energy. The null space of the harmonic energy $S_{\text{harmonic}}(\phi)$ consists of the constant maps $\phi \equiv y$, $y \in N$ (see [11]); that is, all input points in M are mapped to a single point y in N. The property that the harmonic energy penalizes deviations from a constant mapping has severe consequences for the learning task. Namely, if the image of the boundary ∂M is not fixed, then the harmonic energy can always be reduced by contracting the mapping as much as the trade-off between loss and regularizer allows. It is often not easy to know a priori how to fix the image of the boundary boundary ∂M such that no big distortions arise. One example of the negative contraction effects following from this problem can be seen in Figure 6(c) below, another in [26, Figure 4]. It is interesting to note that for the squared geodesic distance loss the learning problem in (4.1) reduces to a classical problem in differential geometry: the task is to find the mean of a set of points on a Riemannian manifold, the so-called Karcher mean [17].

Eells energy. We have shown in the last section that the Eells energy reduces to the classical thin-plate spline energy if input and output manifolds are Euclidean. In this case the null space consists of the linear mappings between input and output space; that is, we are free to fit the data with linear maps, but any deviation from linearity will be penalized. The concept of linearity breaks down in the manifold setting since input and output space have no linear structure.

An interesting question is whether there exists a proper generalization of linear mappings to the case where input and output space are Riemannian manifolds. A key observation towards a natural generalization of the concept of linearity is that linear maps map straight lines to straight lines. Now a straight line between two points in Euclidean space corresponds to a path of shortest length and is thus a geodesic between the two points. In analogy to the Euclidean case we will therefore consider in Riemannian manifolds mappings which map geodesics to geodesics as the proper generalization of linear maps. The following proposition adapted from [11] defines this concept and characterizes the corresponding mappings.

Proposition 4.1 (see [11]). *A map $\phi : M \to N$ is totally geodesic if ϕ maps geodesics of M linearly to geodesics of N; i.e., the image of any geodesic in M is also a geodesic in N though potentially with a different constant speed. The following three properties are equivalent:*

1. *ϕ is totally geodesic;*
2. *ϕ preserves the connection, i.e.,*

$$^N\nabla_{d\phi(X)}d\phi(Y) = d\phi(^M\nabla_X Y),$$

   where $d\phi$ is the differential of ϕ and X, Y are smooth vector fields on M;
3. *$\nabla'_a d\phi^r_b = 0$.*

Proposition 4.1 immediately characterizes the null space of the Eells energy as the set of totally geodesic maps.

Linear maps encode a very simple relation in the data: the relative changes between input and output are the same everywhere. This is the simplest relation that a nontrivial mapping can encode between input and output, and totally geodesic mappings encode the same "linear" relationship even though the input and output manifolds are nonlinear. However, note that as linear maps, totally geodesic maps are *not* necessarily distortion-free, i.e., isometric. However, every distortion-free mapping is totally geodesic. Moreover, given "isometric" training points,

$$d_M(X_i, X_j) = d_N(Y_i, Y_j), \qquad i, j = 1, \dots, k,$$

then among all minimizers of (2.1), there will be an isometry fitting the data points, given that such an isometry exists. With this restriction in mind, one can see the Eells energy also as a measure of distortion of the mapping ϕ. This makes the Eells energy an interesting candidate for a variety of geometric fitting problems, for example, for surface registration as demonstrated in the experimental section.

Despite the similarity of linear and totally geodesic maps it should be noted that there are circumstances in which they show completely different behavior. One important example is discussed in section 8.3.

In contrast to the harmonic energy, the Eells energy does not lead to contraction effects. Imagine the situation of only two given training points in a regression problem from the real line to the sphere. While the solution for the harmonic energy tends to contract and passes only in the limit $\lambda \to 0$ exactly through the points, the solution for the Eells energy yields a geodesic which exactly fits the given training data points for any $\lambda \geq 0$. It also extrapolates "linearly," whereas the harmonic solution which minimizes the change of the prediction function does not extrapolate beyond the first and last training points. These effects are demonstrated in Figures 4 and 6 below.

Biharmonic energy. The null space of the biharmonic energy is a superset of the null space of the Eells energy, since here only the trace of the "Hessian" of ϕ has to vanish, not all its components. Apart from totally geodesic mappings, the null space of the biharmonic energy also contains all stationary maps of the harmonic energy; see Theorem D.2 in Appendix D. Although this sounds reasonable at first, the null space may thus be too big for some applications. This can be seen from a simple example in Euclidean space. Consider the mapping $\phi : \mathbb{R}^2 \to \mathbb{R}$ with $\phi(x_1, x_2) = x_1^2 - x_2^2$, which is clearly nonlinear and intuitively not very smooth. Nevertheless, the biharmonic energy of this mapping is zero. We therefore recommend the Eells energy as a better smoothness measure.

4.2. Difference of biharmonic and Eells energy. One can show (see Theorem 4.2 below) that in Euclidean spaces the biharmonic and the Eells/thin-plate spline energy differ only by a boundary term. In the literature they are therefore often considered as equivalent; see, for example, [10]. However, even in Euclidean space, this is only justified, given that one can guarantee that the first or second derivative of the function one wants to learn vanishes on the boundary of the domain or decays to zero at infinity. Furthermore, if either M or N is non-Euclidean, the two energies are different due to curvature effects, even when one neglects boundary terms. Interestingly, this difference also holds for simple real-valued functions on a non-Euclidean Riemannian manifold M, that is, for $N = \mathbb{R}$. The proof of the following theorem is found in Appendix B.

Theorem 4.2. The biharmonic and Eells energy are related in the following way:

$$S_{biharmonic}(\phi) = S_{Eells}(\phi) + \int_M h_{rs}\, g^{ab}\, g^{cd}\, d\phi_c^r \left(R_{adb}^{M\ e}\, d\phi_e^s - R_{tuv}^{N\ s}\, d\phi_a^t\, d\phi_d^u\, d\phi_b^v \right) dV$$

$$+ \int_{\partial M} N^b h_{rs}\, g^{cd} \left(d\phi_b^r\, \nabla_c' d\phi_d^s - d\phi_c^r\, \nabla_b' d\phi_d^s \right) d\tilde{V},$$

where $R_{adb}^{M\ e}$, $R_{tuv}^{N\ s}$ is the Riemannian curvature tensor of M, N, and $d\tilde{V}$ is the volume form of the boundary ∂M.

4.3. Physical interpretation of intrinsic second-order energies. In [24] classical mechanics is reformulated in terms of differential geometry. Namely, one considers the set of possible configurations of a system to form manifold N. The standard example is rigid body motion which has configuration space $\mathbb{R}^3 \times SO(3)$, that is, position plus orientation. The manifold of configurations is then given a geometric structure by using the kinetic energy as Riemannian metric. Newton's equation for the time-dependent state γ of the physical system, $\gamma : [a, b] \to N$, can then be written as

$$\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t) = \tau(t, \gamma(t)),$$

where τ are the external forces acting upon the system. Noting that it is exactly this acceleration $\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t)$ which is penalized in the biharmonic/Eells energy of curves (3.8), one can interpret the smoothing problem (2.1) for γ as optimizing the trade-off between passing through the set of given points and following free motion as much as possible. Since the acceleration is directly related to the external forces acting on the system, the biharmonic/Eells energy penalizes the amount of external forces which have to act on a physical system to push it along a certain trajectory. For applications like animation or robot control, where a real physical system is lying beneath the learning problem, the biharmonic/Eells energy is thus particularly useful.

5. From intrinsic to extrinsic representation. One can deduce from (3.4) that the representation of the Eells energy in coordinates of M and N is quite complicated and not easily accessible for optimization. Moreover, the use of local coordinate systems introduces the complication that the mapped point $\phi(x)$ can be in different coordinate systems during the optimization.

In this section we show that these difficulties can be elegantly circumvented if M and N are assumed to be isometrically embedded submanifolds in Euclidean spaces $\mathbb{R}^k$ and $\mathbb{R}^l$, respectively. We show that in this case the first- and second-order differential energies presented above have *equivalent* but much simpler forms in terms of the derivatives with respect to the embedding spaces. Expressing the regularization functionals in terms of ambient space also allows the usage of a global coordinate system, which reduces the algorithmic overhead dramatically.

The assumption of the existence of an isometric embedding into Euclidean space is not restrictive. Any compact manifold can be isometrically embedded into Euclidean space $\mathbb{R}^k$ for large enough k; see [28]. For a huge class of manifolds an isometric embedding in Euclidean space is known. Often the manifold is even defined as a constrained set in $\mathbb{R}^k$ or given just as a point cloud in $\mathbb{R}^k$, where in both cases the metric is induced from $\mathbb{R}^k$ and the isometric embedding is trivial.

Below, quantities which are defined on M or N are called intrinsic, whereas quantities related to the embedding spaces $\mathbb{R}^k$ and $\mathbb{R}^l$ are called extrinsic. The goal will be to represent the intrinsic expressions introduced above with simpler computable extrinsic ones. *We stress that in doing this we neither lose the invariance with respect to parametrization nor change the regularizer.*

For simplicity we split the discussion below. We first consider the case where N is a general Riemannian manifold isometrically embedded in $\mathbb{R}^l$, and then the case where M is a
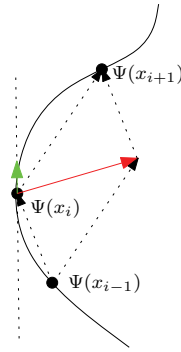
Figure 1. *Comparison of extrinsic and intrinsic second derivative. Suppose $\phi : \mathbb{R} \to N$, N the black curve on the left. Thus, $\Psi : \mathbb{R} \to \mathbb{R}^2$, but $\Psi(x) \in N$. If the images $\Psi(x_i)$ of equidistant points x_i in the input manifold $M = \mathbb{R}$ are also equidistant on the output manifold, then Ψ has no acceleration in terms of N; i.e., its intrinsic second derivative in N should be zero. However, the extrinsic second derivative of Ψ in the ambient space, which is marked red in the figure, is not vanishing in this case. The Eells energy penalizes only the intrinsic acceleration, that is, only the component parallel to the tangent space at $\Psi(x_i)$, the green arrow.*

general manifold embedded in $\mathbb{R}^k$. The proofs of all theorems are in Appendix C.

5.1. Computation of the energies for general output manifolds. Assume that the output manifold N can be embedded isometrically into $\mathbb{R}^l$, and let $i : N \to \mathbb{R}^l$ be the embedding map. Denote by $\Psi : M \to \mathbb{R}^l$ the composition $\Psi = i \circ \phi$. Let z^μ be standard Cartesian coordinates in $\mathbb{R}^l$. Then the differential of Ψ is given as $d\Psi_a^r = \frac{\partial \Psi^\mu}{\partial x^\alpha} dx_a^\alpha \otimes \frac{\partial^r}{\partial z^\mu}$. In order to define derivatives of the differential $d\Psi_a^r$ we need a pull-back connection $\tilde{\nabla} : TM \otimes \Psi^{-1} T\mathbb{R}^l \to \Psi^{-1} T\mathbb{R}^l$ for Ψ; see Appendix A. It is

$$\tilde{\nabla}_{\frac{\partial}{\partial x^\alpha}} \frac{\partial^r}{\partial z^\mu} := {}^{\mathbb{R}^l}\nabla_{d\Psi(\frac{\partial}{\partial x^\alpha})} \frac{\partial^r}{\partial z^\mu} = 0,$$

the second equation due to the flatness of the connection of $\mathbb{R}^l$. Because of this property the expressions for the covariant derivatives of Ψ simplify significantly. However, note that the coordinate vector $\frac{\partial^r}{\partial y^\mu}$ of N has the derivative $\tilde{\nabla}_{\frac{\partial}{\partial x^\alpha}}\big(di(\frac{\partial^r}{\partial y^\mu})\big) = \frac{\partial^2 i^\rho}{\partial y^\nu \partial y^\mu} \frac{\partial \phi^\nu}{\partial x^\alpha} \frac{\partial^r}{\partial z^\rho}$. The following theorem shows how intrinsic expressions in ϕ can be expressed in terms of the extrinsic ones in Ψ.

Theorem 5.1. *The following relations of intrinsic and extrinsic objects hold:*

$$(5.1) \qquad d\phi_a^r = d\Psi_a^r, \qquad \nabla_c' d\phi_a^r = \big(\tilde{\nabla}_c d\Psi_a^r\big)^\top,$$

where $^\top$ denotes the projection onto the tangent space $T_{\Psi(x)} N$ of N.

Theorem 5.1 is visualized in Figure 1. If M is a domain in $\mathbb{R}^m$, it allows us to derive a dramatic simplification of the energy expressions.

Theorem 5.2. *Let $M \subset \mathbb{R}^m$ and x^α be Cartesian coordinates; then*

$$(5.2) \qquad S_{\text{harmonic}}(\Psi) = \int_M \sum_{\mu=1}^l \sum_{\alpha=1}^m \left(\frac{\partial^2 \Psi^\mu}{\partial x^\alpha}\right)^2 dx,$$
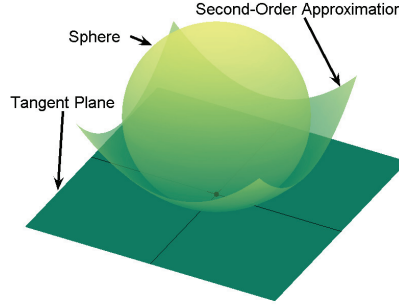
Figure 2. *Second-order approximation of a sphere at the south pole. Note that the principal curvature, also called the* extrinsic curvature, *quantifies how much the manifold bends with respect to the ambient space.*

$$(5.3) \qquad S_{\text{biharmonic}}(\Psi) = \int_M \sum_{\mu=1}^{l} \sum_{\alpha=1}^{m} \left[\left(\frac{\partial^2 \Psi^\mu}{\partial x^\alpha \partial x^\alpha} \right)^\top \right]^2 dx,$$

$$(5.4) \qquad S_{\text{Eells}}(\Psi) = \int_M \sum_{\mu=1}^{l} \sum_{\alpha,\beta=1}^{m} \left[\left(\frac{\partial^2 \Psi^\mu}{\partial x^\alpha \partial x^\beta} \right)^\top \right]^2 dx.$$

5.2. Computation of the energies for general input manifolds. Now assume that the input manifold M is isometrically embedded in $\mathbb{R}^k$. This allows us to construct local second-order parametrizations of M, for which the evaluation of the Christoffel symbols ${}^M\Gamma_{\beta\alpha}^\gamma$ in the second derivative (3.4) is particularly easy.

Proposition 5.3. *Let $x^1, \ldots, x^m$ be the coordinates associated with an orthonormal basis of the tangent space of M at $p \in M$. Then in Cartesian coordinates z of $\mathbb{R}^k$, the manifold can locally be approximated up to second order as*

$$z(x) = \left(x^1, \ldots, x^m, f^{m+1}(x), \ldots, f^k(x) \right),$$

where $f^i(x) = \sum_{\alpha,\beta=1}^{m} \Pi_{\alpha\beta}^i x^\alpha x^\beta$ and $\Pi_{\alpha\beta}^i$ is the second fundamental form of M at p. If M is a hypersurface in $\mathbb{R}^k$ ($k = m + 1$) and if the coordinates x^α are aligned with the principal directions of M at p, then one has $f^k(x) = \sum_{\alpha=1}^{m} \kappa_\alpha (x^\alpha)^2$, where κ_α are the principal curvatures at p.

A visualization of this standard result of differential geometry (see, e.g., [21]) is given in Figure 2. The principal curvature, also called the *extrinsic curvature*, quantifies how much the input manifold bends with respect to the ambient space. Local second-order approximations allow us to compute the second derivative in (3.4) efficiently, as the next proposition shows.

Proposition 5.4. *Given a second-order approximation of M centered at p as in Proposition 5.3, then for coordinates x we have that*

$$g_{\alpha\beta}(0) = \delta_{\alpha\beta}, \qquad {}^M\Gamma_{\beta\gamma}^\alpha(0) = 0.$$

Furthermore, at $p \in M$,

$$(5.5) \qquad \left[\frac{\partial^2 \Psi^\mu}{\partial x^\beta \partial x^\alpha} - \frac{\partial \Psi^\mu}{\partial x^\gamma} {}^M\Gamma^\gamma_{\beta\alpha} \right] = \frac{\partial^2 \Psi^\mu}{\partial x^\beta \partial x^\alpha} = \left[\frac{\partial^2 \Psi^\mu}{\partial z^\beta \partial z^\alpha} + \sum_{r=m+1}^{k} \frac{\partial \Psi^\mu}{\partial z^r} \Pi^r_{\beta\alpha} \right].$$

For a hypersurface M ($k = m + 1$), one has $\Pi^r_{\beta\alpha} = \delta_{\beta\alpha}\kappa_\alpha$ if the coordinates x^α are aligned with the principal directions and κ_α are the principal curvatures of M at p.

Note that (5.5) is not an approximation but the true second derivative of Ψ at p on M. This is due to the following argument: If we allowed for higher-order terms in $f^{m+1}, \ldots, f^k$, we could fit M exactly in a local neighborhood around p. Then x would be coordinates of M and not of its second-order approximation. However, since the computation of Christoffel symbols at p and of (5.5) requires only second derivatives of $f^{m+1}, \ldots, f^k$ at p, we would obtain identical results.

A straightforward consequence of Proposition 5.4 is Corollary 5.5 below, which gives simple extrinsic forms for the Eells and biharmonic energies with manifold-valued input. These expressions are derived by replacing the second partial derivatives in (5.3) and (5.4) with the slightly more complicated expression (5.5). We show only the energy densities here, since coordinates z are different for each point $p \in M$.

Corollary 5.5. *For general input manifolds M and a second-order approximation as in Proposition 5.3, the energy densities of the Eells and biharmonic energies of Ψ at p are given as*

$$(5.6) \qquad \textit{biharmonic:} \qquad \sum_{\mu=1}^{l} \sum_{\alpha=1}^{m} \left[\left(\frac{\partial^2 \Psi^\mu}{\partial z^\alpha \partial z^\alpha} + \sum_{r=m+1}^{k} \frac{\partial \Psi^\mu}{\partial z^r} \Pi^r_{\alpha\alpha} \right)^\top \right]^2,$$

$$(5.7) \qquad \textit{Eells:} \qquad \sum_{\mu=1}^{l} \sum_{\alpha,\beta=1}^{m} \left[\left(\frac{\partial^2 \Psi^\mu}{\partial z^\beta \partial z^\alpha} + \sum_{r=m+1}^{k} \frac{\partial \Psi^\mu}{\partial z^r} \Pi^r_{\beta\alpha} \right)^\top \right]^2.$$

The principal curvatures are known in analytic form for many manifolds. For point cloud data one can estimate them using a local fit with a quadratic function.

5.3. Comparison of intrinsic and extrinsic energies. The expression of the intrinsic second derivative in terms of extrinsic quantities allows us to discuss the differences between our approach, which penalizes only intrinsic variations of the mapping, and the approach recently proposed in [16, 47], where extrinsic variations are penalized. In this section, we will briefly argue why this extrinsic energy has worse properties than our proposed intrinsically defined one, demonstrating the difference for curves $\gamma : M \to N$, $M \subseteq \mathbb{R}$.

Suppose the output manifold N is isometrically embedded in $\mathbb{R}^l$. The extrinsic second-order regularization functional $S_{\mathrm{ex}}(\gamma)$ is given as

$$S_{\mathrm{ex}}(\gamma) = \int_M \|\ddot{\gamma}\|^2 \, dt,$$

where $\ddot{\gamma}$ is the second derivative in $\mathbb{R}^l$. In contrast, the Eells energy $S_{\mathrm{in}}(\gamma)$ reduces for curves to

$$S_{\mathrm{in}}(\gamma) = \int_M \|\nabla_{\dot{\gamma}} \dot{\gamma}\|^2 \, dt.$$
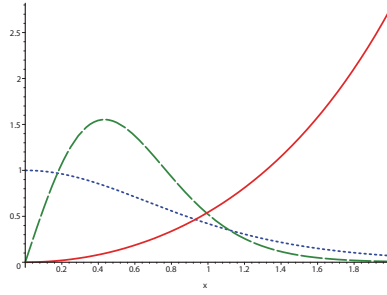
Figure 3. *Example showing that geodesics are in general not minimizers of the extrinsic second-order energy. Solid: the manifold N is given as the graph of a function $f : (0, \infty) \to \mathbb{R}$. Dotted: the curvature of N, that is, the scalar second fundamental form, at $(x, f(x))^T \in N$ as a function of x. Dashed: gradient of the extrinsic energy density $S_{\mathrm{ex}}(\gamma)$ along TN for unit-speed curve $\gamma(t) \in N$. The tangential gradient at $\gamma(t) = (x, f(x))^T$ is plotted as a function of x. While γ is a geodesic in N, the tangential gradient of $S_{\mathrm{ex}}(\gamma)$ does not vanish.*

In both cases one has the constraint $\gamma(x) \in N$ for all $x \in M$. The extrinsic and intrinsic derivative are related via $\ddot{\gamma} = \nabla_{\dot{\gamma}}\dot{\gamma} + \Pi(\dot{\gamma}, \dot{\gamma})$, where $\Pi : TN \times TN \to NN$ is the second fundamental form of N and NN denotes the normal bundle of N; see also Figure 1. That means that the extrinsic energy penalizes the intrinsic tangential acceleration *and* the normal component. We have $\|\ddot{\gamma}\|^2 = \|\nabla_{\dot{\gamma}}\dot{\gamma}\|^2 + \|\Pi(\dot{\gamma}, \dot{\gamma})\|^2$, and therefore

$$S_{\mathrm{ex}}(\gamma) = S_{\mathrm{in}}(\gamma) + \int_M \|\Pi(\dot{\gamma}, \dot{\gamma})\|^2 \, dt.$$

If N has constant extrinsic curvature as, for example, the sphere, then $\|\Pi(\dot{\gamma}, \dot{\gamma})\|^2 = C \|\dot{\gamma}\|^2$, so that the extrinsic energy functional is just a combination of harmonic and Eells energy. For simplicity suppose that we are given only two data points. Using the intrinsic energy, we will find a connecting geodesic as the solution of the learning problem in (2.1), since geodesics have zero energy S_{in}. For the extrinsic energy, the harmonic part of the energy aims to contract the curve; thus the minimum of (2.1) will be a geodesic segment that ends short of the training points, depending on the regularization parameter λ.

While in the special situation above the solutions are at least similar, the extrinsic energy leads to less intuitive solutions in the general case of nonconstant extrinsic curvature. The following simple example shows that geodesic segments are no longer minimizers of the extrinsic energy S_{ex} if the second fundamental form is nonconstant. Yet, they would be global minimizers of the intrinsic energy S_{in}.

Assume now that the output manifold N is the graph of a smooth function $f : (0, \infty) \to \mathbb{R}$, that is, $N = \{x \in \mathbb{R}^2 | x_1 > 0, f(x_1) = x_2\}$ with $f(x) = \cosh(x) - 1$. A unit speed curve in N is given as $\gamma(t) = (\mathrm{arsinh}(t), f(\mathrm{arsinh}(t)))^T$, i.e., $\|\frac{\partial}{\partial t}\gamma(t)\| = 1$. Curves which minimize the extrinsic energy S_{ex} in $\mathbb{R}^2$, equivalent to cubic splines, need to have their fourth derivative orthogonal to N. In Figure 3 we plot the tangential component of $\frac{\partial^4}{\partial t^4}\gamma(t)$, $\langle \dot{\gamma}, \gamma^{(4)} \rangle = 6t/(1 + t^2)^3$, which is clearly nonzero. Thus the geodesic γ is not a minimizer of the extrinsic energy S_{ex}.

6. Implementation. A classic route to solving the variational learning problem (2.1) would be to derive the Euler–Lagrange variational equations and to solve these. We have computed the first variation for all energies in Appendix D. However, this generally leads to systems of fourth-order differential equations which are numerically very difficult to solve. If M is one-dimensional and N is a bi-invariant Lie group, one can reduce such systems to second-order differential equations; see [29, 31]. For general Riemannian manifolds, such reductions are not obvious. Instead, we propose to solve the problem in a way similar to finite element approaches. That is, we directly minimize the optimization problem (2.1) over a parametrized class of functions ϕ. This way, only second derivatives are needed, and furthermore no boundary conditions have to be specified explicitly.

In the following we will explain how to express objective (2.1) in terms of finitely many parameters, and how these can then be optimized efficiently with a pseudo-Newton method. All information about the manifolds that are used in a specific application is made available to the optimization routine through a number of interface functions. An implementation of these interfaces for two standard types of manifolds, namely spheres and point clouds, is described afterward. Since we aim at using the tools from the previous section, we will throughout this section assume that M and N are isometrically embedded in $\mathbb{R}^k$ and $\mathbb{R}^l$, respectively, and the targeted function is thus represented as $\Psi : M \subseteq \mathbb{R}^k \to \mathbb{R}^l$.

6.1. The optimization. Concerning a suitable representation of Ψ in terms of parameters, consider the following arguments. If the output space was Euclidean, then the Euler–Lagrange equations of the different energies derived in Theorem D.2 would be linear differential equations that could elegantly be solved using Green's functions centered at the data points. Optimization problem (2.1) could then be reduced to an optimization in the span of such functions [45]. However, this reasoning is critically dependent on the linear structure of the output space N, and no simple parametric form exists for the minimizer of (2.1) if the output is a general Riemannian manifold, simply because the set of all mappings from M to N is *not* even a vector space.

In the general manifold case, we thus have to resort to some form of discretization. Typical approaches are gridding combined with finite difference approximations for the derivative operators or alternatively finite element methods. Here, we propose using a collocation-like approach, choosing the local polynomials as a flexible function class. Note that the minimizer of (2.1) will almost surely not lie exactly in this set. However, by adding additional polynomial centers we can make the function class more and more flexible and thus approximate the optimal solution more and more closely. The proposed approach also allows computing the required derivatives analytically, without the necessity of a uniform grid on the input space.

Let M be an open subset or submanifold of $\mathbb{R}^k$. Then we parametrize the μth component of mapping $\Psi : \mathbb{R}^k \to \mathbb{R}^l$ as a local polynomial of low order; that is,

$$\Psi^\mu(x) = \frac{\sum_{i=1}^S k_{\sigma_i}(\|\Delta x_i\|) g(\Delta x_i, w_i^\mu)}{\sum_{j=1}^S k_{\sigma_j}(\|\Delta x_j\|)}.$$

Here, $g(\Delta x_i, w_i^\mu)$ is a first- or second-order polynomial in Δx_i with parameters w_i^μ, $\Delta x_i = (x - c_i)$ is the difference between x and the local polynomial centers c_i, and $k_{\sigma_i}(x) = k(r \equiv \frac{x}{\sigma_i})$ is a compactly supported smoothing kernel with bandwidth σ_i. We choose the local polynomial

centers c_i approximately uniformly distributed over M, thereby adapting the function class to the shape of the input manifold M. If we stack all parameters w_i^μ into a single vector w, then Ψ and its partial derivatives are just linear functions of w, which allows computing these values in parallel for many points using simple matrix multiplication.

We compute the energy integral (3.3) as a function of w, by summing up the energy densities at an (approximately) uniformly chosen set of samples from M. The projection onto the tangent space, used in (5.6) and (5.7), and the second-order approximation for computing intrinsic second derivatives, used in (5.6) and (5.7), are manifold-specific and are explained below.

If N is non-Euclidean, which is the case in which we are mostly interested, we need to satisfy the constraints $\Psi(x) \in N$ for $x \in M$ throughout the optimization process. We soften this condition and add it to the objective function as $\gamma \int_M d(\Psi(x), N)^2 dx$, where $d(y, N)$ denotes the Euclidean distance of a point $y \in \mathbb{R}^l$ from the manifold N. We increase the weight γ during the iterative optimization process until all points are within a given prespecified distance of N. As initial solution for Ψ, we compute the free solution, i.e., where N is assumed to be $\mathbb{R}^l$. In this case the problem becomes convex quadratic, since there are no constraints and no location-dependent projections. The iteratively increasing penalization of the distance to the manifold leads to a slow settling of the initial solution towards the target manifold. In contrast to a simple projection of the initial solution onto N, as done in [40], this procedure is much more robust. The projection of Ψ can lead to large distortions which, in turn, can cause the optimization to become numerically unstable or to stop in local minima.

If we allow for $\Psi(x) \notin N$ during the optimization, then we have to declare how the projection of the second derivative of Ψ onto the tangent space is meant and how we deal with the loss term in this case. We propose to determine the projection using the isodistance manifolds $N_{\Psi(x)} = \{y \in \mathbb{R}^l | d(y, N) = d(\Psi(x), N)\}$ of N. For the loss we use the geodesic distance between the projection of $\Psi(X_i)$ onto N and Y_i, that is, $d_N(\mathrm{argmin}_{y \in N} \|\Psi(X_i) - y\|, Y_i)$. These two constructions are sensible, since as the weight γ of the constraint $\gamma \int_M d(\Psi(x), N)^2 dx$ increases, Ψ will approach the manifold N, and both terms converge to the corresponding operations directly executed on the manifold N. The computation of $d(\Psi(x), N)$, the projection onto tangent spaces of isodistance manifolds, and the computations of geodesic distances on N are again manifold-specific and can be found below.

Having expressed all parts of the optimization problem (2.1) in terms of the parameters w, we obtain an unconstrained nonlinear optimization problem $\min_w f(w)$, which we solve using a pseudo-Newton method as follows. For each update we compute the true gradient $\nabla f(w)$, but only an approximation $\tilde{\nabla}^2 f(w)$ of the Hessian, that is, the Hessian of $f(w)$ but without the projection onto the tangent space of N in the Eells energy. We then perform a line search in the direction $-(\tilde{\nabla}^2 f(w))^{-1} \nabla f(w)$ and update w accordingly. Computing only an approximation of the exact Hessian is advantageous for two reasons. First, it is computationally much simpler since no second derivative of the projection operator is required. Second, it adds to the robustness of the algorithm due to the following argument. The Eells energy does not penalize oscillations in the normal direction of the manifold. While such oscillations cannot occur if $\Psi(w) \in N$ is strictly enforced, they can occur during the optimization process where we have relaxed that constraint. Using the approximate Hessian discourages such distorting oscillations; however, we are still guaranteed to minimize the true Eells energy. This can

be seen as follows. The approximate Hessian of the Eells energy is positive semidefinite. If we assume that the training points fix an optimal linear transformation, then the combined approximate Hessian of the whole objective (2.1) is positive definite, and the multiplication of the gradient with the inverse of this matrix just corresponds to a change of the used inner product of the Euclidean embedding space. Finally, note that computation of the descent direction $-(\tilde{\nabla}^2 f(w))^{-1}\nabla f(w)$ can be performed efficiently with sparse methods, since the compact support of the smoothing kernel k implies sparsity of the approximate Hessian $\tilde{\nabla}^2 f(w)$.

6.2. Manifold operations. It remains to describe the required manifold-specific operations. First, we need to be able to project onto the tangent space of the output manifold N and its isodistance manifolds. Second, we need to be able to project from the embedding space of the output manifold $\mathbb{R}^l$ onto N, and third, we require geodesic distances on N. For curved input manifolds M we additionally need the principle curvatures for the intrinsic second derivatives; see Proposition 5.4.

In this section we focus on the two types of manifolds that we used in our experiments, that is, the sphere $\mathcal{S}^{l-1} \subseteq \mathbb{R}^l$ and two-dimensional surfaces in $\mathbb{R}^3$ which are given as point clouds with surface normals. Note that the projection $P^\top$ onto the tangent space of N and its isodistance manifolds can conveniently be performed for any embedded manifold if we have access to a signed distance function η of the manifold N. The projection $P^\top$ at $x \in \mathbb{R}^l$ is then given as $P^\top(x) = 1 - \frac{1}{\|\nabla\eta(x)\|^2}\nabla\eta(x)\nabla\eta(x)^T$.

For the unit sphere $\mathcal{S}^{l-1}$, the signed distance function is simply given as $\eta(x) = 1 - \|x\|$. The projection from the embedding space onto the sphere is trivial, and the geodesic distance is $d(x,y) = \arccos\left(\frac{\langle x,y \rangle}{\|x\|\|y\|}\right)$ for $x, y \in \mathcal{S}^{l-1}$. The principle curvatures are both -1 for all $p \in M$.

For point cloud surfaces in three dimensions, there exist many known methods for constructing signed distance functions, e.g., [33, 41]. Here, we choose a particularly simple approach for computing the signed distance value $\eta(p)$ for some test point $p \in \mathbb{R}^l$: we first search for the closest point to p in the given point cloud, then compute a local second-order approximation there based on the 10 nearest neighbors using least squares, and finally use the distance to this second-order approximation as the desired signed distance function η. The computation of the distance to the local second-order approximation $(x^1, x^2, f(x^1, x^2))$ involves solving third-order equations. However, since we assume that our manifolds are densely sampled, we will always obtain local coordinates (p^1, p^2, p^3) for p with small values for p^1, p^2. Thus, a good approximation to the true distance is to use $\eta(p) = p^3 - f(p^1, p^2)$. The signed distance function so constructed readily allows us to compute the required projections onto the tangent spaces. Furthermore, the same procedure also allows us to determine the closest point on N for a given query point, using just $(p^1, p^2, f(p^1, p^2))$. If the point cloud serves as an input manifold M, the same local second-order approximations are used to obtain the required principal curvatures. What remains is the geodesic distance for point clouds. Either one can use approaches like [20], or alternatively geodesic distances can be computed using the length of a curve which minimizes the harmonic energy and whose endpoints are fixed at the two points of interest [40]. However, since in our surface registration problem we used rather large weights for the loss, $\Psi(X_i)$ and Y_i were always very close on the surface. In
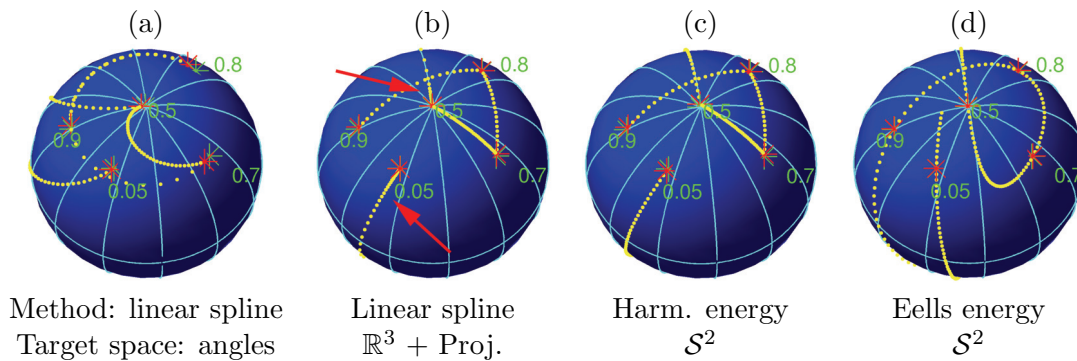
(a)	(b)	(c)	(d)
Method: linear spline	Linear spline	Harm. energy	Eells energy
Target space: angles	$\mathbb{R}^3$ + Proj.	$\mathcal{S}^2$	$\mathcal{S}^2$

Figure 4. *The interval $[0, 1]$ is mapped onto the unit sphere $\mathcal{S}^2$ in three dimensions. Green markers show the given data points $Y_i \in \mathcal{S}^2$, and respective training times $X_i \in [0, 1]$ are given as numbers close by. Red markers indicate $\Psi(X_i)$ for the approximating spline $\Psi : [0, 1] \to \mathcal{S}^2$. Yellow dots mark the Ψ-images of equally spaced points in $[0, 1]$.*

this case the geodesic distance can be well approximated by the Euclidean one, so that for performance reasons we directly used the Euclidean distance.

7. Experiments. We now show some illustrative examples for regression between Riemannian manifolds. The examples show an increasing amount of theoretical and algorithmic complexity. While the first results are similar in flavor to [6, 32], the subsequent examples present novel applications for regression between manifolds.

7.1. Curves on spheres. To understand the basic problems of manifold-valued regression and to get a qualitative idea of the features of our approach, it is helpful to discuss Figure 4 in detail. The aim is here to fit a curve on the sphere $\mathcal{S}^2 \subseteq \mathbb{R}^3$ through six given data points, that is, $\phi : [0, 1] \to \mathcal{S}^2$.

A naive first idea for solving this problem could be to parametrize the surface of the sphere using spherical coordinates and to interpolate the coordinates of the given data points using linear splines. (For visualization purposes we use linear splines corresponding to first-order differential energies here.) This is computationally attractive since the coordinates form a linear space such that the splines can be computed using simple basis function expansions. However, as shown in Figure 4(a), no path can go through the parametrization boundary at $-\pi$ and π, and, moreover, the geometry is heavily distorted by the nonlinear parametrization mapping from $\mathcal{S}^2$ to $(-\pi, \pi) \times (0, \pi)$. Another naive idea, shown in Figure 4(b), is to first compute a linear spline in $\mathbb{R}^3$ and then project it radially onto the sphere. While the trajectory can now surround the sphere, the metric is still distorted through the projection. This can be seen in that the yellow points, which are equally spaced in the input, are not equally spaced in the output; see the locations indicated by the red arrows in Figure 4(b).

Manifold-adapted approaches are much better suited for this regression problem. In Figure 4(c), the harmonic energy (3.2) is used in the learning objective (2.1). Note that the yellow points are now equally spaced between any two data points, up to small distortions resulting from the two-dimensional visualization. However, since the minimizers of the harmonic energy are piecewise geodesic [22], the curve is not differentiable at the data points.
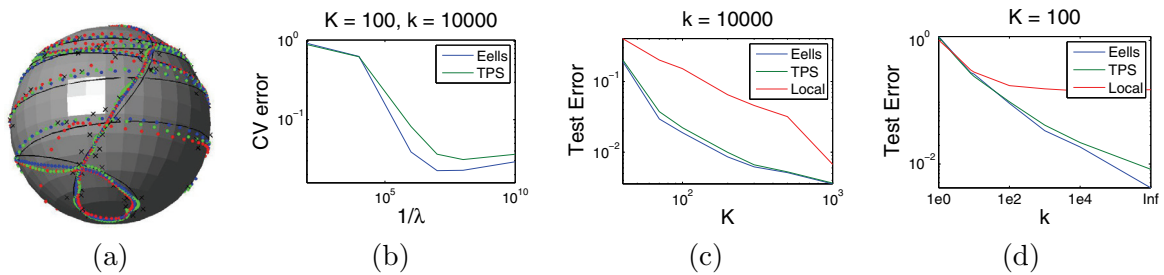
Figure 5. *Regression from $[0,1]$ to the sphere. (a) Noisy data samples (black crosses) of the black ground-truth curve. The blue dots show the estimated curve for our Eells-regularized approach, the green dots depict TPS in $\mathbb{R}^3$ radially projected onto the sphere, and the red dots show results for the local approach of [9]. (b) Cross-validation errors for given sample size K and noise concentration k. Von Mises distributed with $k = 10000$ roughly corresponds to Gaussian noise with standard deviation 0.01. (c) Test errors for different K but fixed k. In all experiments the regularization parameter λ is found using cross-validation. (d) Test errors for different k but fixed K.*

It also does not extend outside of the first/last marker. Using the Eells energy, both these problems are avoided; see Figure 4(d). The curves are smooth and extrapolate linearly or, more precisely, "geodesically."

Turning to quantitative analysis, we expect that a manifold-adapted approach is much better at approximating some unknown curve from noisy observations. As ground-truth we use the curve given in spherical coordinates as $\theta(t) = (40t^2, 1.3\pi t + \pi \sin(\pi t))$. The K training inputs were sampled uniformly from $[0, 1]$, and the outputs perturbed by "additive" noise from the von Mises distribution with concentration parameter k. The von Mises distribution is the maximum entropy distribution on the sphere for fixed mean and variance [23], and thus is the analogue of the Gaussian distribution for spheres. The optimal regularization parameter λ was determined by 10-fold cross-validation; the experiments were repeated 10 times for each training sample size K and noise parameter k.

We compare our framework for nonparametric regression between manifolds with standard cubic smoothing splines in $\mathbb{R}^3$—the equivalent of thin-plate splines (TPS) for one input dimension—projected radially on the sphere, and also with the local manifold-valued Nadaraya–Watson estimator of [9] using Gaussian weights, where the bandwidth parameter is also determined by cross-validation. As can be seen in Figure 5, our globally regularized approach performs significantly better than [9] for this task. One can observe in Figure 5(a) that, even in places where the estimated curve of [9] follows the ground-truth relatively closely, the spacing between points varies greatly. These sampling-dependent speed changes, which are not seen in the ground truth curve, cannot be avoided without a global smoothness prior such as, for example, the Eells energy. The Eells approach also outperforms the projected TPS method, particularly for small sample sizes and reasonable noise levels. For a fixed noise level of $k = 10000$ a paired t-test showed that our reduction in test error is statistically significant at level $\alpha = 5\%$ for the sample sizes $K = 70, 200, 300, 500$. Clearly, as the curve is very densely sampled for high K, both approaches perform similarly, since the problem then is essentially local and the manifold is locally linear. For small sample sizes, however, i.e., for situations where the a priori information is more important, the TPS method is outper-
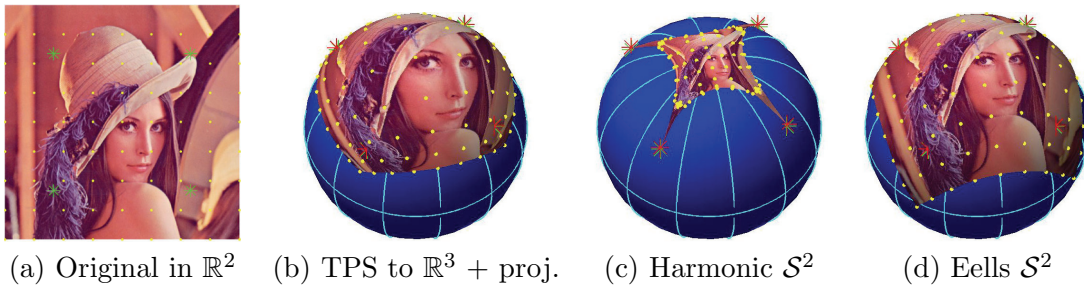
(a) Original in $\mathbb{R}^2$ (b) TPS to $\mathbb{R}^3$ + proj.   (c) Harmonic $\mathcal{S}^2$ (d) Eells $\mathcal{S}^2$

Figure 6. *The Lena image* (a) *is used to visualize a mapping from the unit square in $\mathbb{R}^2$ to the unit sphere $\mathcal{S}^2$ in $\mathbb{R}^3$. Green markers show the given data point pairs, and red stars on $\mathcal{S}^2$ denote positions of the input markers in $\mathbb{R}^2$ mapped to the sphere by the approximating spline.* (b) *shows a TPS mapping from $\mathbb{R}^2$ to $\mathbb{R}^3$ projected onto $\mathcal{S}^2$.*

formed by the proposed Eells-regularized approach, as we are using an application-adapted smoothness functional.

7.2. Two-dimensional patches on the sphere. As in the last section, we here demonstrate qualitative differences between projected TPS, the harmonic energy, and the Eells energy solution, but now for a two-dimensional input manifold $M = [0,1]^2 \subset \mathbb{R}^2$. Such a setup is common for many geometric modelling tasks such as surface parametrization, remeshing, or texture mapping.

In Figure 6(b), we first compute a TPS solution in $\mathbb{R}^3$, which in this case yields a plane cutting through the four given markers, and then project it radially onto the sphere. This results in an extreme fish-eye distortion. In Figure 6(c), the results are shown for the variational setting using the harmonic energy. The harmonic energy is commonly used in geometric modelling, e.g., [50], although mostly in targeting linear spaces. The mapped image does not fill the convex hull of the training points, and we observe an undesirable contraction of the image. This is why the harmonic energy is traditionally used only for input domains without boundary, or when the output boundary can be fixed a priori. While there exist methods to alleviate this problem [50], a theoretically clean way would be to use the proposed Eells energy as a regularizer; see Figure 6(d). Since the Eells energy does not try to minimize the distances between the points, but rather the variation of the distances, it is much less prone to contraction of the image. Moreover, it extrapolates nicely out of the convex hull of the training points. Furthermore, the distortion-minimizing property of the Eells energy can be observed here nicely. While it is not possible to exactly map all geodesics in the input to geodesics in the output, the Eells regularized approach performs much better than the projected TPS solution in Figure 6(b).

7.3. Fingerprints. One task in fingerprint processing is estimating the direction of the fingerprint's contour lines; see Figure 7. In some regions where the contrast is high the direction can be reliably estimated based on image gradients. The goal is now to predict the directions in all regions based on the available estimated direction data. This can be formulated as a regression problem, where the input domain M is the image $\Omega \in \mathbb{R}^2$ and the output the set of line directions which equals real projective space $\mathbb{RP}_1$. $\mathbb{RP}_1$ is up to a global scaling factor isometric isomorphic to S^1. Thus we use this simpler representation as output
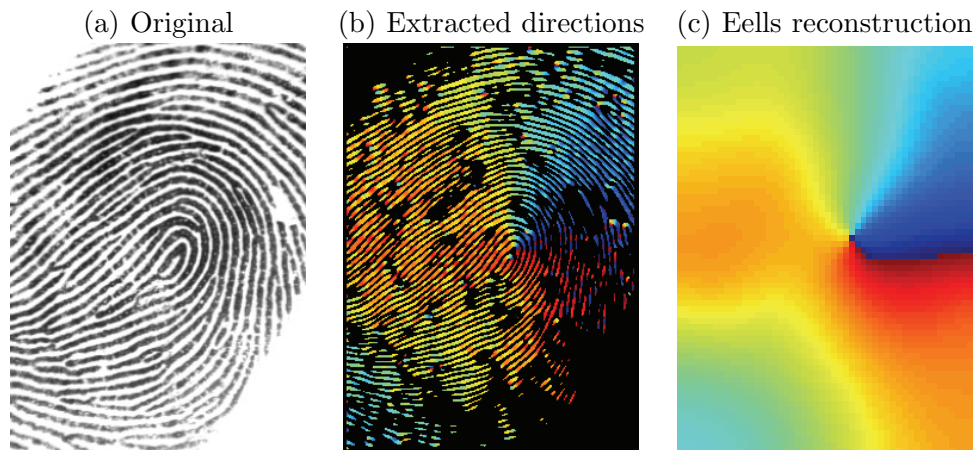
(a) Original   (b) Extracted directions   (c) Eells reconstruction

Figure 7. *Smoothing and hole filling fingerprints.* (a) *Original image* I. (b) *Extracted directions color-coded. For the black regions the direction cannot be extracted reliably, since the image gradient $\|\nabla I\|$ is too small.* (c) *Eells energy–based reconstruction. Note that dark red and dark blue colors actually code for the same direction.*

space N.

As training and validation data, we use directions computed from the polar representation of the gradients of the smoothed original image; see Figure 7(a). In low contrast regions with radial gradient component smaller than 50 (image values ranging from 0 to 255) the orientation extraction was deemed too unreliable to be considered, yielding the holey ground-truth image shown in (b). We computed our reconstruction shown in (c) using Eells energy–based regression. As training data, we used a subset of the nonblack pixels from (b), extracted on a grid with 10 pixel spacing. Nonblack off-grid pixels in (b) were used for validation purposes. The regularization parameter λ was chosen by 5-fold cross-validation.

The result shown in Figure 7(c) is a nicely smoothed version of the original noisy and holey dataset (b). The validation error is 14.8°. We compare our approach with the diffusion method in [34], which minimizes an approximation of the harmonic energy. As opposed to [34], our method is also able to fill holes in the dataset. Thus, when comparing on the same training data as above, we had to initialize the diffusion method with the full direction image that also covered the regions marked unreliable in (b). With an optimally selected data term, the best validation error we could achieve for their method was 17.7°, significantly higher than for our Eells energy–based approach.

In total we did experiments with 47 fingerprint images taken from [43]. Selected images had at least 20% of the pixels with gradient norm above the threshold. This guaranteed not only sensible smoothing, but, more importantly, reliable validation. We obtained the following results:

	Method of [34]	Eells-based approach
Validation error	25° ± 6°	22° ± 6°

The relative improvement of our method is on average 12%. This is significant under a t-test at level 5% (P-value 2×10^{-4}).

7.4. Surface/head correspondence. Computing correspondence between the surfaces of different but similar objects, such as human heads, is a central problem in shape processing. A dense correspondence map, that is, an assignment of all points of one head to the anatomically equivalent points on the other head, allows one to perform morphing [37] or to build linear object models [5], also known as active appearance models [8], which are flexible tools for computer graphics as well as computer vision. Though the problem is well studied, it remains challenging and is still actively investigated. Most approaches minimize a functional that consists of a local similarity measure and a smoothness functional or regularizer for the overall mapping. Motivated by the fact that the Eells energy favors "linear" mappings, we propose to use it as regularizer for correspondence maps between surface manifolds. For testing and highlighting the role of this "prior" independent of the choice of local similarity measure, we formulate the dense correspondence problem as a nonparametric regression problem between manifolds where 55 point correspondences on characteristic local texture or shape features are given. (Only on the forehead we fix some less well-defined markers, to determine a relevant length-scale.)

It is in general difficult to evaluate correspondences numerically, since for different heads anatomical equivalence is not easily specified. Here, we have used a subset of the head database of [5] and considered the correspondence developed there as our ground-truth. These correspondences are known to be perceptually highly plausible. We took the average head of one part of the database and registered it to the other 10 faces, using the mean distance to the correspondence of [5] as an error score. Apart from the average deviation over the whole head, we also show results for an interior region (see Figure 8(g)), for which the correspondence given by [5] is known to be more exact compared to other regions as, for example, around the ear or below the chin.

We compared our approach against [37] and a TPS-like approach. The TPS method represents the initial solution of our approach, that is, a mapping into $\mathbb{R}^3$ minimizing the TPS energy (3.7), which is then projected onto the heads' surfaces. [37] uses a volume-deformation based approach that directly finds smooth mappings from surface to surface, without the need of projection, but their regularizer does not take into account the true distances along the surface. We did not compare against the local regression approach of [9], since their approach requires computing a large number of geodesics in each iteration. This is computationally prohibitive on point clouds. In order to obtain a sufficiently flexible, yet not too high-dimensional, function set for our implementation, we place polynomial centers c_i on all markers points and also use a coarse, approximately uniform sampling of the other parts of the manifold. Free parameters, that is, the regularization parameter λ and the density of additional polynomial centers, were chosen by 10-fold cross-validation for our and the TPS methods, and by manual inspection for the approach of [37].

One computed correspondence example is shown in Figure 8, and the average over all 10 test heads is summarized in the table below.

	TPS	Eells	[37]
Mean error for the full head (mm)	2.90	2.16	2.15
Mean error for the interior (mm)	1.49	1.17	1.36

The manifold-adapted Eells approach performs much better than the TPS method, especially in regions of high curvature such as around the nose; see the error heatmaps in Figure 8.
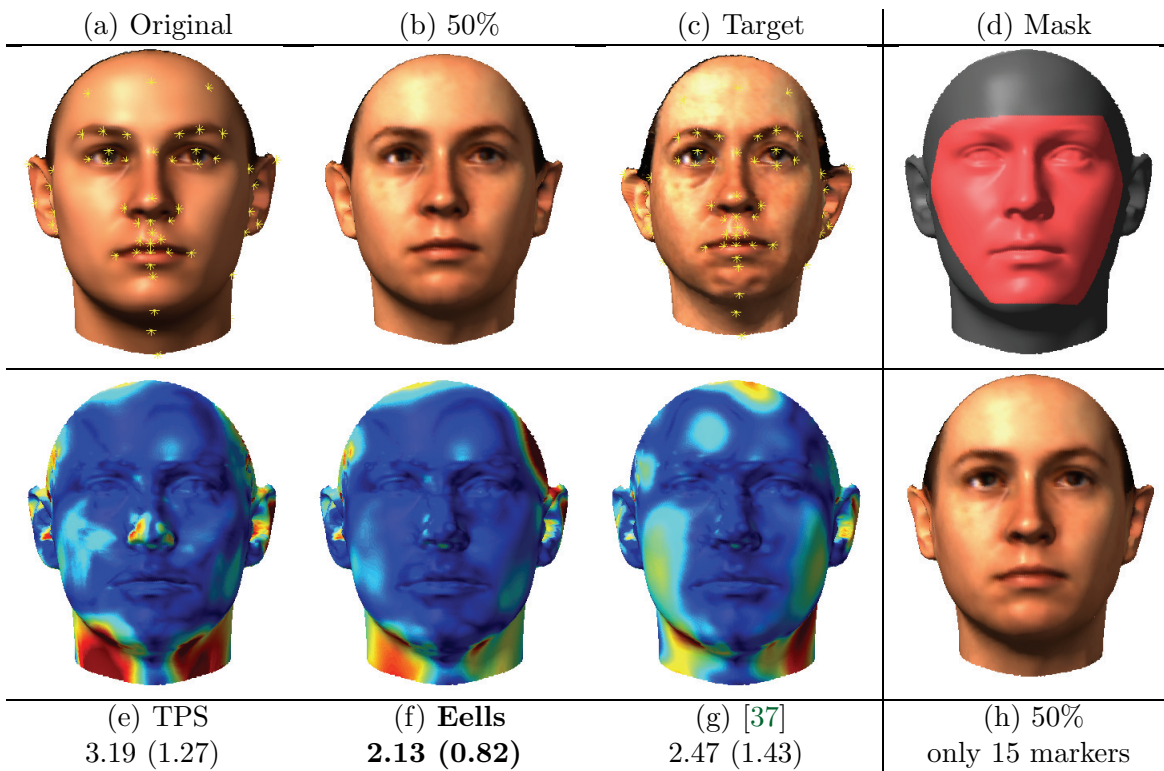
(a) Original	(b) 50%	(c) Target	(d) Mask

(e) TPS	(f) **Eells**	(g) [37]	(h) 50%
3.19 (1.27)	**2.13 (0.82)**	2.47 (1.43)	only 15 markers

Figure 8. *Correspondence computation from the original head in* (a) *to the target head in* (c) *with* 55 *markers (yellow crosses). A resulting* 50% *morph using our method is shown in* (b). *Distance of the computed correspondence from the correspondence of* [5] *is color-coded in* (e)–(g) *for different methods. The numbers below give the average distance in mm over the whole head, and in brackets the average over an interior region (red area in* (d)*). Using our method with only* 15 *markers (see* (h)*) still yields visually plausible morphing results.*

Compared to [37], our method finds a smoother, more plausible solution, particularly on large textureless areas such as the forehead or the cheeks.

We repeated the experiment with only very few markers (15). While the alignment of small texture details then becomes troublesome, which negatively affects numeric evaluation, the overall visual impression of our method is still fairly good; see Figure 8(h). This shows once more that the Eells energy is a suitable prior for mappings between three-dimensional object surfaces.

8. Further topics in manifold-valued learning. In this final section we want to point out structural differences of regression between Riemannian manifolds and standard multivariate regression. The results derived here are rather preliminary, and the goal of this section is more to identify interesting open problems than to provide an already fully developed solution.

8.1. Function spaces. The regularized risk minimization objective (2.1) is minimized over all smooth mappings $C^\infty(M, N)$. It is a classical problem in variational analysis that this space is not sufficient to guarantee the existence of a minimizer, since it is not complete. For Euclidean output, one therefore introduces the Sobolev space $W^{s,2}(M, \mathbb{R}^l)$ as the completion

of $C^s(M, \mathbb{R}^l)$ with respect to the norm

$$\|\phi\|_s^2 = \sum_{\mu=1}^{l} \sum_{r=0}^{s} \int_M \|\nabla_1 \cdots \nabla_r \phi^\mu\|^2 \; dV.$$

The functions in $W^{s,2}(M, \mathbb{R}^l)$ need not be in $C^s(M, \mathbb{R}^l)$, but at least it is known that a (weak) minimizer of the energy functional exists in $W^{s,2}(M, \mathbb{R}^l)$. For example, consider linear splines: the minimizers of (2.1) using the harmonic energy in $W^{1,2}(\mathbb{R}, \mathbb{R})$ are piecewise linear but not differentiable at the data points $\phi(X_i)$. Under strong assumptions a similar result has been derived for linear splines in manifolds without extending the theory of Sobolev-spaces to manifold-valued output in [22]. However, a general approach which uses fewer assumptions and which is also valid for higher-dimensional input will most likely require such an extension.

One problem with extending Sobolev theory to manifolds is that, if the output manifold is non-Euclidean, any class of functions from M to N can *not* be a vector space due to the missing linear structure of N. This is problematic, as the vector spaces structure is typically one of the first abstractions introduced in derivations of Sobolev space, and it is also heavily used in the proofs of existence and uniqueness of minimizers of (2.1) for Euclidean output. Discarding this structure thus requires one to make fundamental changes. Instead of a vector space structure, the space of admissible functions should rather be thought of as an infinite-dimensional manifold where the tangent spaces have Hilbert space structure. Results for such a generalization in the case of harmonic energy have been obtained in [15, 48]. However, the generalization to higher-order regularization functionals is an open problem.

Another problem occurs already for Euclidean outputs. If $\dim(M) \geq 2s$, point evaluation functionals on $W^{s,2}(M, \mathbb{R}^l)$ are not necessarily continuous; see Sobolev's embedding theorem [2]. This implies that point interpolation or approximation with regularization of order s is not well defined in this case. Since Euclidean output is only a special case of the more general manifold-valued setup and since Riemannian manifolds are locally linear, we assume that the problem similarly transfers to the manifold case. In the Euclidean setting, a classic route to obtaining a well-defined learning objective (2.1) is to resort to higher-order regularization, since for any input dimension m sufficiently large s guarantees continuity of point evaluation. Higher-order regularization is feasible in the Euclidean-output case, since the optimal solution is known to be a sum of Green's functions centered at the data points [45], and since these Green's functions can be computed analytically [49]. In the general manifold-valued output setting, however, the solution needs to be computed numerically via discretization, and higher-order regularization leads to ever-more-complicated expressions for the derivatives. Thus, this direction seems infeasible at the moment. Note that for second-order functionals such as the Eells and biharmonic energy the critical condition is $\dim(M) \leq 3$. In all our experiments, this condition was fulfilled, and we thus assume that a sensible well-defined minimizer of (2.1) exists. For higher-dimensional input we plan to investigate in the future whether smoothing of the loss or switching to the $W^{s,p}$ norm, $p > 2$, offers a solution to the described problem.

8.2. Homotopy and consistency. In the following we will explore the nontrivial topological structure of manifold-valued mappings.

Definition 8.1. *Two continuous mappings ϕ_1, ϕ_2 from M to N are said to be homotopic if there exists a continuous mapping $\Psi : M \times [0, 1] \to N$ with $\Psi(x, 0) = \phi_1(x)$, $\Psi(x, 1) = \phi_2(x)$.*

Homotopy defines an equivalence relation on $C(M, N)$. We denote the set of resulting equivalence classes, the so-called homotopy classes, by $[M, N]$. One says that $[M, N]$ is trivial if it consists of just the homotopy class of the constant map. It is easy to see that $[M, \mathbb{R}^l]$, that is, the homotopy class of mappings considered in manifold learning, is trivial. However, for the manifold-valued output this is generally not the case, with interesting theoretical as well as practical implications.

Typically, the regularized empirical risk minimization problem is solved using a descent-type algorithm which continuously deforms the current mapping ϕ. This implies that the homotopy class is preserved during optimization, and thus the homotopy class of the final solution is determined by the initial mapping. Theoretically, one could just search for the optimum in all homotopy classes. This is, however, practically not possible, since, e.g., $[\mathcal{S}^1, \mathcal{S}^1]$ is isomorphic to the set of integers—the number of cycles around the circle. Thus, consistent estimation practically requires that the initial solution be already in the same homotopy class as the "true" solution, i.e., the Bayes optimal solution $\gamma^* = \arg\min_{\gamma \text{ measurable}} \mathbb{E}_{Y,X}\, d^2(\gamma(X), Y)$.

The following theorem provides a first step towards a consistent training procedure for manifold-valued mappings, where $[M, N]$ is nontrivial.

Theorem 8.2. *Given K training points $(X_i, Y_i) \in \mathcal{S}^1 \times \mathcal{S}^1$, let h be the maximal geodesic nearest neighbor distance of $\{X_i\}_{i=1}^K$. Let the Bayes optimal solution γ^* be smooth and the sampling distribution be deterministic, i.e., $\mathrm{P}(\gamma^*(X) \neq Y) = 0$. Then for $\left\|\dot{\gamma^*}\right\| \leq L$ and $h < \frac{\pi}{L}$ the piecewise geodesic interpolant of the training data is in the same homotopy class as γ^*.*

Proof. Let X_i and X_j be nearest neighbors in $\mathcal{S}^1$. We have $\int_{X_i}^{X_j} \left\|\dot{\gamma^*}\right\| dt \leq L\, d_{S_1}(X_i, X_j) \leq L\, h$. With $L\, h < \pi$ we know that γ^* can have made no cycle around $\mathcal{S}^1$ between X_i and X_j. Moreover, the length of the shortest path between Y_i and Y_j is also bounded by $L\, h < \pi$. Thus the geodesic $\hat{\gamma}$ interpolating (X_i, Y_i) and (X_j, Y_j) is homotopic to the segment of $\gamma^*|_{X_i}^{X_j}$. Since this holds for any neighboring points of the training data, the whole curves γ^* and $\hat{\gamma}$ are homotopic. ∎

The theorem can be extended to nondeterministic problems where $\mathrm{P}(Y|X)$ is sufficiently concentrated, and to the setting where $(X_i, Y_i)_{i=1}^K$ is a random sample from P on $\mathcal{S}^1 \times \mathcal{S}^1$. The generalization of this result to more general domains is nontrivial and is an interesting problem for future research.

8.3. Capacity of totally geodesic maps. In section 4.1 we have argued that totally geodesic maps share important geometric properties with linear maps in Euclidean spaces. We now examine whether this also holds in a statistical perspective.

In statistics, linear maps are favorable since they are small capacity approximators. This means that it is unlikely that one can fit a random training set exactly with linear maps, which renders regression with linear maps relatively robust against overfitting [36]. Since totally geodesic maps are identical to linear maps if one has Euclidean input and output, their capacity is small, too, at least for some, most likely for many combinations of input-output manifolds. However, the following example demonstrates that this is not always true. There exists a nontrivial connection between the topology of the output manifold and the capacity of the associated totally geodesic maps.

Consider again mappings from $M = \mathcal{S}^1$ to $N = \mathcal{S}^1$. In standard angular coordinates, all

totally geodesic maps in this setting are of the form $\phi_a(x) = \text{mod}(a\,x + b)$ for $a \in \mathbb{N}$ and $b \in [0, 2\pi)$. The following theorem, which is a classical result in number theory, shows that this set of mappings can fit any given set of training points arbitrarily well and thus has infinite capacity.

Theorem 8.3 (see [1, p. 154]). *Let $(X_i, Y_i) \in \mathcal{S}^1 \times \mathcal{S}^1$, $i = 1, \ldots, K$, be the training data. Then there exists for any set of training data and any $\varepsilon > 0$ an $a \in \mathbb{N}$ such that*

$$\max_{i=1,\ldots,K} d(\phi_a(X_i), Y_i) \le \varepsilon.$$

Since totally geodesic mappings are not penalized by the Eells energy, the solution of regularized empirical risk minimization in (2.1) is always given by the geodesic ϕ_a, which obviously overfits the training data. However, note that the integer a, which corresponds to the number of cycles around the circle of ϕ_a, (empirically) grows exponentially with the number of data points. This is the reason why we did not encounter this phenomenon in the implementation of [40]. The above phenomenon still holds if the input space is the real line or a closed interval. At least for regression into $\mathcal{S}^1$ this example suggests that the null-space of both the Eells and the biharmonic energies of manifold-valued mappings is already too large to be useful. Since for the harmonic energy one has $S_{\text{harm}}(\phi_a) = 2\pi a$, this suggests that from a theoretical point of view one should use in this case either the harmonic energy or a combination of harmonic and a second-order energy.

9. Conclusion. This paper has presented a universal theoretically sound framework for regression between two Riemannian manifolds based on regularized empirical risk minimization. The discussed differential energies are dependent only on the geometry of the input and output manifold, but not on their respective parametric representation. We have derived an intuitively desirable property of the proposed Eells energy, namely that it favors the so-called totally geodesic maps, a suitable generalization of linear maps. Our implementation and our experimental results have further supported the benefits of using a truly manifold-adapted approach and especially the Eells energy.

Throughout the paper we tried to convey that the problem of manifold-valued regression is far from being a trivial generalization of the Euclidean case, and there remain many challenging and interesting open questions in the mathematical and statistical analysis of this problem. On the practical side, an interesting question is whether there exists a compact but flexible representation for general mappings between Riemannian manifolds. Since our implementation is based on discretization, it is so far limited to low-dimensional input spaces. However, for many statistical problems higher-dimensional input would be desirable, requiring a more compact function representation. In Euclidean space this is typically done with sparse basis function expansions. However, since manifold-valued output does not allow for the addition of functions, this route cannot be taken here. The construction of compact yet flexible representations for mappings between general Riemannian manifolds thus remains an important open point.

Appendix A. The pull-back connection, its curvature, and Green's theorem. This section is a self-contained review of the basic ingredients of connections and the curvature of vector bundles. With the exception of the extension of the Green's theorem to the tensor product connection, the material can be found in [11].

Let M be a smooth, connected, orientable Riemannian manifold. Let V be a smooth vector bundle over M of finite rank with base projection $\pi : V \to M$. We denote by $C(V)$ the vector space of smooth sections of V, i.e., of smooth maps $\sigma : M \to V$ such that $\pi \circ \sigma = \mathbf{1}_M$. Let V and W be two vector bundles over M; then we define the following:

- V^* is the dual bundle of V,
- $V \oplus W$ is the direct sum of V and W,
- $V \otimes W$ is the tensor product of V and W,
- $\otimes^p V$ is the pth tensor power of V,
- $\wedge^p V$ is the pth exterior power of V (completely antisymmetric),
- $\odot^p V$ is the pth tensor power of V (completely symmetric).

An important concept for manifold-valued mappings is the pull-back bundle $\phi^{-1}W$.

Definition A.1. If $\phi : M \to N$ and W is a vector bundle over N, then $\phi^{-1}W$ denotes the pull-back bundle, whose fiber over $x \in M$ is $W_{\phi(x)}$, the fiber of W over $\phi(x)$.

Next we define the Riemannian metric and the connection on vector bundles.

Definition A.2. A Riemannian metric on a vector bundle V is a section a in $C(V^* \odot V^*)$ which induces on each fiber a positive definite inner product. Let $\sigma, \rho \in C(V)$; then we use $\langle \sigma, \rho \rangle := a(\sigma, \rho)$.

As in the case of the tangent bundle, one can introduce the musical isomorphisms to define maps $V \to V^*$ and $V^* \to V$. One can also define a Riemannian metric on the pull-back bundle. Let $\phi : M \to N$, and W be a vector bundle over N with metric b. We can identify $\sigma, \rho \in (\phi^{-1}W)_x$ with $\sigma, \rho \in W_{\phi(x)}$ and thereby define $\langle \sigma, \rho \rangle_b$.

Definition A.3. A linear connection on vector bundle V over M is a bilinear map ∇ on spaces of sections,

$$\nabla : C(TM) \times C(V) \to C(V), \qquad (X, \sigma) \mapsto \nabla_X \sigma,$$

where $X \in C(TM)$ and $\sigma \in C(V)$ such that for $f \in C(M)$ we have

- $\nabla_{fX}\sigma = f\nabla_X \sigma,$
- $\nabla_X(f\sigma) = X(f)\,\sigma + f\,\nabla_X \sigma.$

Since ∇ is linear in its first argument, we write in abstract index notation $X^a \nabla_a \sigma^{t_1,\ldots,t_s}_{b_1,\ldots,b_r}$ for a (s, r) vector bundle V.

Definition A.4. Let $^V\nabla$ and $^W\nabla$ be connections on V and W.

1. *The dual connection on V^* is defined by*

   (A.1) $\qquad \theta \in C(V^*),\ \sigma \in C(V); \qquad (\nabla_X \theta)(\sigma) = X(\theta(\sigma)) - \theta(\nabla_X \sigma).$

2. *The direct sum connection on $V \oplus W$ is defined as*

   (A.2) $\qquad \sigma \in C(V), \lambda \in C(W); \qquad \nabla_X(\sigma \oplus \lambda) = {}^V\nabla_X \sigma \oplus {}^W\nabla_X \lambda.$

3. *The tensor product connection on $V \otimes W$ is defined as*

   (A.3) $\qquad \sigma \in C(V), \lambda \in C(W); \qquad \nabla_X(\sigma \otimes \lambda) = {}^V\nabla_X \sigma \otimes \lambda + \sigma \otimes {}^W\nabla_X \lambda.$

The following definition of the pull-back connection is the central key to the definition of energy functionals for manifold-valued mappings.

Definition A.5. *For smooth map* $\phi : M \to N$ *and vector bundle* W *over* N *with connection* $^W\nabla$, *the* pull-back *or* induced connection *on* $\phi^{-1}W$ *is defined as the connection* ∇' *on* $\phi^{-1}W$ *such that for each* $x \in M$, $X \in T_x M$, *and* $\lambda \in C(W)$,

$$\nabla'_X \left(\phi^* \lambda\right) = \phi^* \left(^W\nabla_{d\phi(X)} \lambda\right),$$

where $d\phi : T_x M \to T_{\phi(x)} N$ *is the push-forward or differential of* ϕ *and* $\phi^* \lambda = \lambda \circ \phi \in C(\phi^{-1}W)$. *In abstract index notation,*

$$\nabla'_a \lambda(\phi(x)) = d\phi_a^r\ {}^W\nabla_r \lambda\big|_{\phi(x)}.$$

This definition, which formally applies only to elements $\phi^* \lambda \in \phi^{-1}W$ derived from $\lambda \in C(W)$, can be uniquely extended to all elements of $\phi^{-1}W$ by using the defining properties of a connection [11].

Definition A.6. *A* Riemannian structure *on bundle* V *is a pair* (∇, a), *where* a *is a Riemannian metric,* ∇ *is a connection, and* $\nabla a = 0$, *where* ∇a *is defined using the tensor product connection in* (A.3).

The condition $\nabla a = 0$ means that for all $X \in C(TM)$, $\sigma, \omega \in C(V)$ we have

$$X \langle \sigma, \omega \rangle = \langle \nabla_X \sigma, \omega \rangle + \langle \sigma, \nabla_X \omega \rangle ;$$

i.e., the connection is compatible with the inner product. Let $(^V\nabla, a)$ and $(^W\nabla, b)$ be Riemannian structures on V and W; then one can check that the direct sum, the tensor product, and the pull-back connection are again Riemannian structures.

Definition A.7. *The* curvature tensor of a connection *is the map* $R : C(TM) \wedge C(TM) \otimes C(V) \to C(V)$ *defined by*

$$R(X,Y)\sigma = \nabla_X \nabla_Y \sigma - \nabla_Y \nabla_X \sigma - \nabla_{[X,Y]} \sigma = -R(Y,X)\sigma.$$

Lemma A.8. *Let* R^V *and* R^W *be the curvature tensors of* V *and* W. *Then,*
- *for* V^*, $(R(X,Y)\theta)(\sigma) = -\theta(R(X,Y)\sigma)$ *for all* $X, Y \in C(TM)$, $\theta \in C(V^*)$, *and* $\sigma \in C(V)$;
- *for* $V \oplus W$, $R(X,Y)(\sigma \oplus \lambda) = R^V(X,Y)\sigma \oplus R^W(X,Y)\lambda$, *where* $\lambda \in C(W)$;
- *for* $V \otimes W$, $R(X,Y)(\sigma \otimes \lambda) = R^V(X,Y)\sigma \otimes \lambda + \sigma \otimes R^W(X,Y)\lambda$;
- *for* $\phi^{-1}W$, $R_x(X,Y)\rho(x) = R^W_{\phi(x)}(d\phi(X), d\phi(Y))\rho(x)$, *where* $\rho \in C(\phi^{-1}W)$.

From here on, we consider only connections derived from the Levi–Civita connections on tangent bundles on M and N. In particular, for smooth map $\phi : M \to N$ we repeatedly consider on $\phi^{-1}TN$ the pull-back connection ∇' of the Levi–Civita connection on N. For mixed tensor $T_a^r \in T^*M \otimes \phi^{-1}TN$ we apply the tensor product connection by using $^M\nabla$ for T^*M and ∇' for $\phi^{-1}TN$. With some abuse of notation we use the same symbol ∇' for all such tensor product connections on $\otimes^k TM \otimes^l T^*M \otimes \phi^{-1}TN$, and refer to them as the *pull-back connections* for all these bundles. The following recipe for a covariant derivative of mixed tensor T can be generalized straightforwardly:

$$\nabla'_b T_a^r = \nabla'_b (T_\alpha^\mu \, dx_a^\alpha \otimes \partial_\mu^r)$$
$$:= \left(^M\nabla_b T_\alpha^\mu\right) dx_a^\alpha \otimes \partial_\mu^r + T_\alpha^\mu \left(^M\nabla_b dx_a^\alpha\right) \otimes \partial_\mu^r + T_\alpha^\mu \, dx_a^\alpha \otimes \left(\nabla'_b \partial_\mu^r\right).$$

As an example consider the differential $d\phi_a^r : T_x M \to T_{\phi(x)} N$,

$$d\phi_a^r(x) = \frac{\partial \phi^\mu}{\partial x^\alpha} dx_a^\alpha \Big|_x \otimes \frac{\partial^r}{\partial y^\mu} \Big|_{\phi(x)} = {}^M\nabla_a \phi^\mu \Big|_x \otimes \frac{\partial^r}{\partial y^\mu} \Big|_{\phi(x)}.$$

We obtain the coordinate expression for $\nabla_b' d\phi_a^r$ as

$$\nabla_b' d\phi_a^r = {}^M\nabla_b {}^M\nabla_a \phi^\mu \otimes \frac{\partial^r}{\partial y^\mu} + {}^M\nabla_a \phi^\mu \otimes \nabla_b' \frac{\partial^r}{\partial y^\mu}$$

$$\text{(A.4)} \qquad = \left[\frac{\partial^2 \phi^\mu}{\partial x^\beta \partial x^\alpha} + \frac{\partial \phi^\mu}{\partial x^\gamma} {}^M\Gamma_{\beta\alpha}^\gamma + \frac{\partial \phi^\rho}{\partial x^\alpha} \frac{\partial \phi^\nu}{\partial x^\beta} {}^N\Gamma_{\nu\rho}^\mu \right] dx_b^\beta \otimes dx_a^\alpha \otimes \frac{\partial^r}{\partial y^\mu}.$$

One can read off that $\nabla_b' d\phi_a^r = \nabla_a' d\phi_b^r$, because the Levi–Civita connections on M and N are symmetric, that is, because ${}^M\Gamma_{\beta\alpha}^\gamma = {}^M\Gamma_{\alpha\beta}^\gamma$ and ${}^N\Gamma_{\nu\rho}^\mu = {}^N\Gamma_{\rho\nu}^\mu$. With this in mind, we can show the following lemma, which will be useful later on.

Lemma A.9. *Let $\phi : M \to N$ and $X, Y \in C(TM)$; then we have*

$$\nabla_X'(d\phi(Y)) - \nabla_Y'(d\phi(X)) = d\phi([X,Y]),$$

where $[X,Y]$ is the Lie-bracket.

Proof. We have

$$X^b \nabla_b'(d\phi_a^r Y^a) - Y^b \nabla_b'(d\phi_a^r X^a)$$
$$= d\phi_a^r(X^b \, {}^M\nabla_b Y^a - Y^b \, {}^M\nabla_b X^a) + X^b Y^a [\nabla_b' d\phi_a^r - \nabla_a' d\phi_b^r] = d\phi_a^r [X,Y]^a,$$

where we used in the first step the definition of the pull-back connection for tensor product spaces, and in the second step $\nabla_b' d\phi_a^r = \nabla_a' d\phi_b^r$. ∎

We generalize Green's theorem to the case of the pull-back connection as follows.

Lemma A.10. *Let $T \in C(\otimes^{p+1} T^* M \otimes \phi^{-1} TN)$ and $S \in C(\otimes^p T^* M \otimes \phi^{-1} TN)$. Then, with ∇' being the pull-back connection, we have*

$$\int_M \langle T, \nabla' S \rangle = \int_{\partial M} \langle T, N \otimes S \rangle - \int_M \langle \text{trace}_g \nabla' T, S \rangle,$$

where N is the covector associated with the normal vector at ∂M and the trace is taken with respect to the first two indices. In abstract index notation we obtain

$$\int_M g^{ac_0} g^{b_1 c_1} \cdots g^{b_p c_p} \, h_{rs} \, T_{c_0 \cdots c_p}^r \, \nabla_a' S_{b_1 \cdots b_p}^s$$
$$= \int_{\partial M} g^{ac_0} g^{b_1 c_1} \cdots g^{b_p c_p} \, h_{rs} \, T_{c_0 \cdots c_p}^r \, N_a S_{b_1 \cdots b_p}^s$$
$$- \int_M g^{ac_0} g^{b_1 c_1} \cdots g^{b_p c_p} \, h_{rs} \left(\nabla_a' T_{c_0 \cdots c_p}^r \right) S_{b_1 \cdots b_p}^s.$$

Proof. We show the result for $T \in C(T^* M \otimes \phi^{-1} TN)$ and $S \in C(\phi^{-1} TN)$ using explicit coordinates. The extension to higher tensor powers in $T^* M$ is then straightforward. With $\nabla_a' S^s = \nabla_a'(S^\nu \frac{\partial^s}{\partial y^\nu}) = ({}^M\nabla_a S^\nu) \frac{\partial^s}{\partial y^\nu} + S^\nu \nabla_a' \frac{\partial^s}{\partial y^\nu}$ we can write the part of the covariant derivative associated with the pull-back connection explicitly,

$$\text{(A.5)} \qquad \int_M g^{ab} h_{rs} T_b^r \nabla_a' S^s = \int_M g^{ab} h_{\mu\nu} T_b^\mu [{}^M\nabla_a S^\nu + S^\rho \, {}^N\Gamma_{\rho\omega}^\nu d\phi_a^\omega].$$

Furthermore, we have

$$\int_M g^{ab} h_{\mu\nu} T_b^{\mu}\, {}^M\nabla_a S^{\nu}$$
$$= \int_M g^{ab}\, {}^M\nabla_a (h_{\mu\nu} T_b^{\mu} S^{\nu}) - \int_M g^{ab}\, ({}^M\nabla_a h_{\mu\nu}) T_b^{\mu} S^{\nu} - \int_M g^{ab} h_{\mu\nu} ({}^M\nabla_a T_b^{\mu}) S^{\nu}$$
$$= \int_{\partial M} N^b h_{rs} T_b^r S^s - \int_M g^{ab} \frac{\partial h_{\mu\nu}}{\partial y^{\rho}}\, d\phi_a^{\rho} T_b^{\mu} S^{\nu} - \int_M g^{ab}\, h_{\mu\nu}\, S^{\nu}\, {}^M\nabla_a T_b^{\mu},$$

where we use standard Green's theorem [21] in the second step. With $\frac{\partial h_{\mu\nu}}{\partial y^{\rho}} = h_{\nu\omega}\, {}^N\Gamma_{\rho\mu}^{\omega} + h_{\mu\omega}\, {}^N\Gamma_{\rho\nu}^{\omega}$ we get

$$\int_M g^{ab} \frac{\partial h_{\mu\nu}}{\partial y^{\rho}}\, d\phi_a^{\rho}\, T_b^{\mu}\, S^{\nu} = \int_M g^{ab} [h_{\nu\omega}\, {}^N\Gamma_{\rho\mu}^{\omega} + h_{\mu\omega}\, {}^N\Gamma_{\rho\nu}^{\omega}]\, d\phi_a^{\rho}\, T_b^{\mu}\, S^{\nu}.$$

Plugging the expression for $\int_M g^{ab} h_{\mu\nu} T_b^{\mu}\, {}^M\nabla_a S^{\nu}$ into (A.5), we obtain

$$\int_M g^{ab}\, h_{rs}\, T_b^r\, \nabla_a' S^s = \int_{\partial M} N^b\, h_{rs}\, T_b^r\, S^s - \int_M g^{ab}\, h_{\mu\nu}\, S^{\nu}\, [{}^M\nabla_a T_b^{\mu} + T_b^{\omega}\, {}^N\Gamma_{\omega\alpha}^{\mu}\, d\phi_a^{\alpha}]$$
$$= \int_{\partial M} N^b\, h_{rs}\, T_b^r\, S^s - \int_M g^{ab}\, h_{rs}\, S^s\, \nabla_a' T_b^r. \qquad \blacksquare$$

Appendix B. Proof for section 4.

Proof of Theorem 4.2. We write the difference between the biharmonic and Eells energies as a divergence of a vector field on M plus some curvature terms. We define

$$F_b = h_{rs}\, g^{cd} \left(d\phi_b^r\, \nabla_c' d\phi_d^s - d\phi_c^r\, \nabla_b' d\phi_d^s \right).$$

We have

(B.1)
$$g^{ab}\, \nabla_a' F_b = h_{rs}\, g^{ab}\, g^{cd} \left(\nabla_a' d\phi_b^r\, \nabla_c' d\phi_d^s + d\phi_b^r\, \nabla_a' \nabla_c' d\phi_d^s \right.$$
$$\left. - \nabla_a' d\phi_c^r \nabla_b' d\phi_d^s - d\phi_c^r \nabla_a' \nabla_b' d\phi_d^s \right).$$

The divergence contains the energy densities of the Eells and biharmonic energies plus two other terms. The last term in (B.1) can be rewritten using $\nabla_b' d\phi_d^s = \nabla_d' \phi_b^s$ and

$$\nabla_a' \nabla_d' d\phi_b^s = \nabla_d' \nabla_a' d\phi_b^s - R^M{}_{adb}{}^e\, d\phi_e^s + R^N{}_{tuv}{}^s\, d\phi_a^t\, d\phi_d^u\, d\phi_b^v,$$

where we have used Lemma A.8 in Appendix A for elements in $T^*M \otimes \phi^{-1}TN$ like $d\phi_b^s$. The first term of this new expansion and the second term in (B.1) cancel. Applying the extended Green's theorem, Lemma A.10, we obtain the desired result. $\qquad \blacksquare$

Appendix C. Extrinsic representation of the pull-back connection and proofs for section 5.
Here, we compute a representation of the pull-back connection for manifolds N which are isometrically embedded in Euclidean space.

Definition C.1. *Let ∇' be the connection pull-back via ϕ and $\tilde{\nabla}$ the connection pull-back via $\Psi = i \circ \phi$. The pull-back second fundamental from $\Pi' : TM \otimes \phi^{-1}TN \to \phi^{-1}(TN)^{\perp}$ is defined via*

$$X^a \tilde{\nabla}_a S^r = X^a \nabla'_a S^r + X^a \Pi'^r_{as} S^s.$$

Lemma C.2. *The* pull-back second fundamental form $\Pi'^e_{ab} : TM \otimes \phi^{-1}TN \to \phi^{-1}(TN)^{\perp}$ *can be computed as*

$$\Pi'^r_{as} = d\phi^u_a \, {}^N\Pi^r_{us},$$

where ${}^N\Pi$ is the second fundamental form of N.

Proof. We have $\Psi = i \circ \phi$. With $\frac{\partial^r}{\partial y^\mu} = \frac{\partial i^\alpha}{\partial y^\mu} \frac{\partial^r}{\partial z^\alpha}$ we get

$$d\Psi^r_a = \frac{\partial \Psi^\alpha}{\partial x^\beta} dx^\beta_a \otimes \frac{\partial^r}{\partial z^\alpha} = \frac{\partial i^\alpha}{\partial y^\mu} \frac{\partial \phi^\mu}{\partial x^\beta} dx^\beta_a \otimes \frac{\partial^r}{\partial z^\alpha} = \frac{\partial \phi^\mu}{\partial x^\beta} dx^\beta_a \otimes \frac{\partial^r}{\partial y^\mu} = d\phi^r_a.$$

For $S^r \in \phi^{-1}TN$ one then obtains

$$\tilde{\nabla}_a S^r = d\Psi^s_a \, {}^{\mathbb{R}^l}\nabla_s S^r = d\phi^s_a \left[{}^N\nabla_s S^r + {}^N\Pi^r_{su} S^u \right] = \nabla'_a S^r + d\phi^s_a \, {}^N\Pi^r_{su} S^u.$$

The generalization to covariant derivatives of $\otimes^m T^*M \otimes \phi^{-1}TN$ follows easily. ∎

Now the proofs of section 5 can be derived as follows.

Proof of Theorem 5.1. For $d\Psi^r_a = d\phi^r_{\underline{a}}$, see the proof of Lemma C.2. Moreover, a decomposition of $\tilde{\nabla}_b d\Psi^r_a$ as in Definition C.1 is $\tilde{\nabla}_b d\Psi^r_a = \nabla'_b d\Psi^r_a + \Pi'^r_{bs} d\Psi^s_a$, where $\nabla'_b d\Psi^r_a = \nabla'_b d\phi^r_a \in TM^* \otimes TM^* \otimes \phi^{-1}TN$ and $\Pi'^r_{bs} d\Psi^s_a = d\phi^u_b \, {}^N\Pi^r_{us} d\phi^s_a \in TM^* \otimes TM^* \otimes \phi^{-1}(TN)^{\perp}$ due to Lemma C.2. This shows the result. ∎

Proof of Theorem 5.2. Application of Theorem 5.1 together with $g^{ab} = \delta^{ab}$ for Cartesian coordinates and $\tilde{\nabla}_b d\Psi^r_a = \frac{\partial^2 \Psi^\mu}{\partial x^\alpha \partial x^\beta} dx^\beta_b \otimes dx^\beta_a \otimes \frac{\partial^r}{\partial z^\mu}$ yields the results. ∎

Proof of Proposition 5.4. The function $i : \mathbb{R}^m \to \mathbb{R}^k$ defined here as $(x^1, \ldots, x^m) \mapsto i(x) = (x^1, \ldots, x^m, f^{m+1}(x), \ldots, f^k(x))$ can be seen as the embedding of the second-order approximation of M into $\mathbb{R}^k$. The induced metric is given as

$$g_{\alpha\beta} = \sum_{r=1}^k \frac{\partial i^r}{\partial x^\alpha} \frac{\partial i^r}{\partial x^\beta} = \begin{cases} 1 + \sum_{r=m+1}^k \left(\frac{\partial f^r}{\partial x^\alpha} \right)^2 & \text{if } \alpha = \beta, \\ \sum_{r=m+1}^k \frac{\partial f^r}{\partial x^\alpha} \frac{\partial f^r}{\partial x^\beta} & \text{if } \alpha \neq \beta. \end{cases}$$

Since the functions f^r are all quadratic in coordinates x^α, we immediately see that $g_{\alpha\beta}(0) = \delta_{\alpha\beta}$ and $\frac{\partial g_{\alpha\beta}}{\partial x^\gamma}(0) = 0$, and thus $\Gamma^\gamma_{\alpha\beta} = \frac{1}{2} g^{\gamma\rho} (\partial_\alpha g_{\beta\rho} + \partial_\beta g_{\alpha\rho} - \partial_\rho g_{\alpha\beta}) = 0$ at the origin [21, p. 70]. Finally, we have

$$\frac{\partial^2 \Psi^\mu}{\partial x^\beta \partial x^\alpha} = \frac{\partial^2 \Psi^\mu}{\partial z^r \partial z^u} \frac{\partial z^r}{\partial x^\alpha} \frac{\partial z^u}{\partial x^\beta} + \frac{\partial \Psi^\mu}{\partial z^r} \frac{\partial^2 z^r}{\partial x^\alpha \partial x^\beta}$$

and

$$\frac{\partial z^r}{\partial x^\alpha} = \begin{cases} 1 & \text{if } r = \alpha, \\ 0 & \text{if } r \leq m \text{ and } r \neq \alpha, \\ \frac{\partial f^r}{\partial x^\alpha} & \text{if } r > m, \end{cases} \qquad \frac{\partial^2 z^r}{\partial x^\beta \partial x^\alpha} = \begin{cases} 0 & \text{if } r \leq m, \\ \Pi^r_{\alpha\beta} & \text{if } r > m. \end{cases}$$

The result in (5.5) follows. ■

Appendix D. Variation of the harmonic, biharmonic, and Eells energy. In this section, we derive necessary conditions for the minimizers of the energy functionals, that is, the Euler–Lagrange equations. The variation of the energy functionals is based on the extended Green's theorem, Lemma A.10, and the commutator formula from Lemma D.1 for the exchange of derivatives of the induced connection.

Letting $I = (-\varepsilon, \varepsilon)$, we denote by $\phi(t, x)$, $t \in I$, a variation of the mapping ϕ such that $\phi(0, x) = \phi(x)$, and by $T(M \times I)$ the tangent space of the product manifold $M \times I$. Note that $T(M \times I)$ is isomorphic to $TM \oplus TI$. The product metric is given as $g = g_{TM} \oplus g_{TI}$ and is block-diagonal in any local coordinate system. This implies that also all other structures on the product manifold like Christoffel-symbols or the curvature tensor have this block-diagonal structure.

Lemma D.1. *Let ∇' be the pull-back connection on $T^*(M \times I) \otimes \phi^{-1}TN$. Then*

$$(D.1) \qquad \frac{\partial^a}{\partial t} \nabla'_a d\phi^r_b = \nabla'_b \frac{\partial \phi^r}{\partial t} = \nabla'_b \left(d\phi^r_a \frac{\partial^a}{\partial t} \right),$$

$$(D.2) \qquad \frac{\partial^c}{\partial t} \nabla'_c \nabla'_a d\phi^r_b = \nabla'_a \nabla'_b \frac{\partial \phi^r}{\partial t} + R^N_{suv}{}^r \frac{\partial \phi^s}{\partial t} d\phi^u_a \, d\phi^v_b.$$

Proof. Since $\frac{\partial}{\partial_t}$ and $\frac{\partial}{\partial_{x_i}}$ are coordinate vectors, we have $[\frac{\partial}{\partial_t}, \frac{\partial}{\partial_{x_i}}] = 0$. Moreover, the tensor product of the pull-back connection of $\phi^{-1}TN$ and $T^*(M \times I)$ is compatible with the Riemannian structure on $T^*(M \times I) \otimes \phi^{-1}_t TN$. Using the result of Lemma A.9 with $Y^a = \frac{\partial^a}{\partial t} \in T(M \times I)$, we have

$$X^b \nabla'_b \left(d\phi^r_a \frac{\partial^a}{\partial t} \right) - \frac{\partial^a}{\partial t} \nabla'_a \left(d\phi^r_b X^b \right) = 0.$$

With $\frac{\partial^b}{\partial t} \nabla'_b (d\phi^r_a X^a) = X^a \frac{\partial^b}{\partial t} \nabla'_b d\phi^r_c + d\phi^r_c \frac{\partial^b}{\partial t} \nabla'_b X^a$ and $\frac{\partial^a}{\partial t} \nabla'_a X^b = 0$ (X^b is a vector field on M and does not change with t), we obtain

$$(D.3) \qquad \nabla'_b \left(d\phi^r_a \frac{\partial^a}{\partial_t} \right) = \nabla'_b \frac{\partial \phi^r}{\partial t} = \frac{\partial^a}{\partial t} \nabla'_b d\phi^r_a = \frac{\partial^a}{\partial t} \nabla'_a d\phi^r_b,$$

where the last equality follows by the symmetry of $\nabla'_d d\phi^r_c$. Taking the derivative of (D.3), we get

$$\nabla'_a \nabla'_b \frac{\partial \phi^r}{\partial t} = \left(\nabla'_a \frac{\partial^c}{\partial t} \right) \nabla'_c d\phi^r_b + \frac{\partial^c}{\partial t} \nabla'_a \nabla'_c d\phi^r_b = \frac{\partial^c}{\partial t} \nabla'_a \nabla'_c d\phi^r_b,$$

where we have used that $\left(\nabla'_a \frac{\partial^c}{\partial t} \right)\big|_{TM} = 0$. We will now exchange the order of the derivatives in front of $d\phi^r_b$ using the definition of the curvature tensor for objects of type $T^*(M \times I) \otimes \phi^{-1}TN$,

$$\nabla'_c \nabla'_a d\phi^r_b = \nabla'_a \nabla'_c d\phi^r_b - R^{M \times I\, d}_{cab} d\phi^r_d + R^N_{suv}{}^r d\phi^s_c d\phi^u_a d\phi^b_u,$$

where we have used that the curvature tensor of $M \times I$ is the direct sum of the curvature of M and the curvature of I, which is zero. Moreover, we have, due to the block-diagonal structure

of the curvature tensor, $\frac{\partial^c}{\partial t} R_{cab}^{M \times I\ d} d\phi_d^r = 0$. Thus,

$$\frac{\partial^c}{\partial t} \nabla_c' \nabla_a' d\phi_b^r = \nabla_a' \nabla_b' \frac{\partial \phi^r}{\partial t} + R_{suv}^{N\ \ r} \frac{\partial \phi^s}{\partial t} d\phi_a^u d\phi_u^v,$$

where we have used the previous result. ∎

The previous lemma basically tells us that the time derivative commutes with the pull-back connection. But the "Hessian" does not commute with the time derivative, and one gets an additional curvature term.

Theorem D.2. *Let* $I = (-\varepsilon, \varepsilon)$ *and* $\phi(t, x) : I \times M \to N$ *be a variation of the mapping* $\phi = \phi(0, x)$, *and let* $W^b = \frac{\partial}{\partial t} \phi_t^b \big|_{t=0}$ *be the variational vector field at* $t = 0$. *The variation of the* harmonic energy *is given as*

$$\frac{1}{2} \frac{d}{dt} S_{\mathrm{harmonic}}(\phi_t) \Big|_{t=0} = - \int_M g^{ac} h_{rs} W^r \nabla_c' d\phi_a^s \, dV + \int_{\partial M} h_{rs} N^c W^r d\phi_c^s d\tilde{V}.$$

The variation of the Eells energy *is given as*

$$\frac{1}{2} \frac{d}{dt} S_{\mathrm{Eells}}(\phi_t) \Big|_{t=0} = \int_M g^{ab} g^{cd} h_{rs} W^r \left[\nabla_c' \nabla_a' \nabla_b' d\phi_d^s + R_{twv}^{N\ \ s} d\phi_a^v \, d\phi_c^w \, \nabla_b' d\phi_d^t \right] dV$$
$$+ \int_{\partial M} h_{rs} \, g^{ab} \, N^c \left[\nabla_a' W^r \nabla_c' d\phi_b^s - W^r \, \nabla_a' \nabla_b' d\phi_c^s \right] d\tilde{V}.$$

The variation of the biharmonic energy *is given as*

$$\frac{1}{2} \frac{d}{dt} S_{\mathrm{biharmonic}}(\phi_t) \Big|_{t=0} = \int_M g^{ac} g^{bd} h_{rs} W^r \left[\nabla_c' \nabla_a' \nabla_b' d\phi_d^s + R_{twv}^{N\ \ s} d\phi_a^v \, d\phi_c^w \, \nabla_b' d\phi_d^t \right] dV$$
$$+ \int_{\partial M} h_{rs} \, g^{ab} \, N^c \left[\nabla_c' W^r \nabla_b' d\phi_a^s - W^r \, \nabla_c' \nabla_b' d\phi_a^s \right] d\tilde{V},$$

where $d\tilde{V}$ *is the volume element of the boundary* ∂M, $R_{uvw}^{N\ \ s}$ *is the curvature tensor of* N, *and* N^a *is the normal vector field at* ∂M.

Proof. For the harmonic energy we get, using Lemma D.1 and the extended Green's theorem of Lemma A.10,

$$\frac{1}{2} \frac{d}{dt} S_{\mathrm{harmonic}}(\phi_t) \Big|_{t=0} = \int_M g^{ab} \, h_{rs} \, \nabla_a' W^r \, d\phi_b^s \, dV(x)$$
$$= \int_{\partial M} W^r \, h_{rs} \, N^b \, d\phi_b^s \, d\tilde{V} - \int_M W^r \, h_{rs} \, g^{ab} \, \nabla_a' d\phi_b^s.$$

For the Eells energy we use the commutator of Theorem D.1 and obtain

$$\frac{1}{2} \frac{d}{dt} S_{\mathrm{Eells}}(\phi_t) = \int_M g^{ab} \, g^{cd} \, h_{rs} \, \nabla_a' \nabla_c' \frac{\partial \phi^r}{\partial t} \, \nabla_b'(d\phi_t)_d^s dV$$
$$+ \int_M g^{ab} \, g^{cd} \, h_{rs} \, R_{uvw}^{N\ \ r} \frac{\partial \phi_t^u}{\partial t} (d\phi_t)_a^v (d\phi_t)_c^w \, \nabla_b'(d\phi_t)_d^s \, dV.$$

Using $\nabla'_b(d\phi_t)^s_d\big|_{t=0} = \nabla'_b d\phi^s_d$ and twice the extended Green's theorem yields

$$
\begin{aligned}
\frac{1}{2}\frac{d}{dt}&S_{\text{Eells}}(\phi_t)\Big|_{t=0}\\
&= \int_M g^{ab}\,g^{cd}\,h_{rs}\,\nabla'_a\nabla'_c W^r\,\nabla'_b d\phi^s_d\,dV + \int_M g^{ab}\,g^{cd}\,h_{rs}\,R^N_{uvw}{}^r\,W^u\,d\phi^v_a\,d\phi^w_c\,\nabla'_b d\phi^s_d\,dV\\
&= \int_{\partial M} N^b\,g^{cd}\,h_{rs}\,\nabla'_c W^r\,\nabla'_b d\phi^s_d\,d\tilde V - \int_{\partial M} g^{ab}\,N^d\,h_{rs}\,W^r\,\nabla'_a\nabla'_b d\phi^s_d\,d\tilde V\\
&\quad + \int_M g^{ab}\,g^{cd}\,h_{rs}\,W^r\,\nabla'_c\nabla'_a\nabla'_b d\phi^s_d\,dV + \int_M g^{ab}\,g^{cd}\,h_{rs}\,R^N_{uvw}{}^r\,W^u\,d\phi^v_a\,d\phi^w_c\,\nabla'_b d\phi^s_d\,dV.
\end{aligned}
$$

The result follows, noting that $R_{uvws} = R_{wsuv}$. The variation of the biharmonic energy can be derived analogously. ∎

A necessary condition for a minimizer of the energy $S(\phi)$ is that $\frac{d}{dt}S(\phi_t)\big|_{t=0} = 0$ for all vector fields $W = \frac{\partial\phi}{\partial t}$.

Corollary D.3. *For all points in the interior of $M\backslash\{X_1,\ldots,X_K\}$ the minimizer $\phi : M \to N$ of the learning objective* (2.1) *satisfies for the*

harmonic energy: $\qquad\qquad\qquad\qquad\qquad\qquad g^{ac}\nabla'_c d\phi^r_a = 0,$

biharmonic energy: $\qquad g^{ac}g^{bd}\Big[\nabla'_c\nabla'_a\nabla'_b d\phi^r_d + R^N_{twv}{}^r\,d\phi^v_a\,d\phi^w_c\,\nabla'_b d\phi^t_d\Big] = 0,$

Eells energy: $\qquad g^{ab}g^{cd}\Big[\nabla'_c\nabla'_a\nabla'_b d\phi^r_d + R^N_{twv}{}^r\,d\phi^v_a\,d\phi^w_c\,\nabla'_b d\phi^t_d\Big] = 0.$

The following are natural boundary conditions at ∂M for the

harmonic energy: $\qquad\qquad\qquad N^c d\phi^r_c = 0,$

biharmonic energy: $\qquad\qquad g^{ab}\,\nabla'_b d\phi^r_a = 0,\;\; N^c g^{ab}\nabla'_c\nabla'_b d\phi^r_a = 0,$

Eells energy: $\qquad\qquad\quad N^c\nabla'_c d\phi^r_b = 0,\;\; N^c g^{ab}\nabla'_a\nabla'_b d\phi^r_c = 0.$

The boundary conditions for the biharmonic and Eells energies are sufficient but not necessary for a minimizer. That means they guarantee that the sum of the two boundary terms in the variation vanishes; however, they are not the weakest possible conditions on ϕ. The given boundary conditions are nevertheless "natural" in the sense that both ϕ and its derivative can be arbitrarily chosen on the boundary.

REFERENCES

[1] T. APOSTOL, *Modular Functions and Dirichlet Series in Number Theory*, Springer, New York, 1990.
[2] T. AUBIN, *Some Nonlinear Problems in Riemannian Geometry*, Springer, Berlin, 1998.
[3] M. BELKIN AND P. NIYOGI, *Semi-supervised learning on manifolds*, Machine Learning, 56 (2004), pp. 209–239.
[4] R. BHATTACHARYA AND V. PATRANGENARU, *Large sample theory of intrinsic and extrinsic sample means on manifolds* I, Ann. Statist., 31 (2003), pp. 1–29.
[5] V. BLANZ AND T. VETTER, *A morphable model for the synthesis of 3d faces*, in Proceedings of the SIGGRAPH'99 Conference, Los Angeles, 1999, ACM Press, New York, pp. 187–194.

[6] S. R. Buss and J. P. Fillmore, *Spherical averages and applications to spherical splines and interpolation*, ACM Trans. Graphics, 20 (2001), pp. 95–126.

[7] M. Camarinha, F. Silvia Leite, and P. Crouch, *Splines of class C^k on non-Euclidean spaces*, IMA J. Math. Control Inform., 12 (1995), pp. 399–410.

[8] T. Cootes, G. Edwards, and C. Taylor, *Active appearance models*, IEEE Trans. Pattern Anal. Machine Intell., 23 (2001), pp. 681–685.

[9] B. C. Davis, P. T. Fletcher, E. Bullitt, and S. Joshi, *Population shape regression from random design data*, in Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV), Rio de Janeiro, Brazil, 2007, pp. 1–7.

[10] T. Duchamp and W. Stuetzle, *Spline smoothing on surfaces*, J. Comput. Graphical Statist., 12 (2003), pp. 354–381.

[11] J. Eells and L. Lemaire, *Selected Topics in Harmonic Maps*, AMS, Providence, RI, 1983.

[12] J. Eells and J. H. Sampson, *Harmonic mappings of Riemannian manifolds*, Amer. J. Math., 86 (1964), pp. 109–160.

[13] N. I. Fisher, T. Lewis, and B. J. J. Embleton, *Statistical Analysis of Spherical Data*, Cambridge University Press, Cambridge, UK, 1993.

[14] S. A. Gabriel and J. T. Kajiya, *Spline Interpolation in Curved Manifolds*, manuscript, 1985.

[15] F. Hélein and J. C. Wood, *Harmonic maps*, in Handbook on Global Analysis, Elsevier, New York, 2008, pp. 417–491.

[16] M. Hofer and H. Pottmann, *Energy-minimizing splines in manifolds*, ACM Trans. Graphics, 23 (2004), pp. 284–293.

[17] H. Karcher, *Riemannian center of mass and mollifier smoothing*, Comm. Pure Appl. Math., 30 (1977), pp. 509–541.

[18] W. Kendall, *Probability, convexity, and harmonic maps with small image*. I. *Uniqueness and fine existence*, Proc. London Math. Soc. (3), 61 (1990), pp. 371–406.

[19] M. Kilian, N. Mitra, and H. Pottmann, *Geometric modeling in shape space*, ACM Trans. Graphics (TOG), 26 (2007), article 64.

[20] R. Kimmel and J. Sethian, *Computing geodesic paths on manifolds*, Proc. Natl. Acad. Sci. USA, 95 (1998), pp. 8431–8435.

[21] J. M. Lee, *Riemannian Manifolds—An Introduction to Curvature*, Springer, New York, 1997.

[22] L. Machado, F. S. Leite, and K. Hüper, *Riemannian means as solutions of variational problems*, LMS J. Comput. Math., 9 (2006), pp. 86–103.

[23] K. Mardia and P. Jupp, *Directional Statistics*, Wiley, New York, 2000.

[24] J. Marsden and T. Ratiu, *Introduction to Mechanics and Symmetry*, Springer, New York, 1999.

[25] D. Massonnet, M. Rossi, C. Carmona, F. Adragna, G. Peltzer, K. Feigl, and T. Rabaute, *The displacement field of the Landers earthquake mapped by radar interferometry*, Nature, 364 (1993), pp. 138–142.

[26] F. Mémoli, G. Sapiro, and S. Osher, *Solving variational problems and partial differential equations mapping into general target manifolds*, J. Comput. Phys., 195 (2004), pp. 263–292.

[27] S. Montaldo and C. Oniciuc, *A short survey on biharmonic maps between Riemannian manifolds*, Rev. Un. Mat. Argentina, 47 (2006), pp. 1–22.

[28] J. Nash, *The imbedding problem for Riemannian manifolds*, Ann. Math., 63 (1956), pp. 20–63.

[29] L. Noakes, *Null cubics and Lie quadratics*, J. Math. Phys., 44 (2003), pp. 1436–1448.

[30] L. Noakes, G. Heinzinger, and B. Paden, *Cubic splines on curved spaces*, IMA J. Math. Control Inform., 6 (1989), pp. 465–473.

[31] L. Noakes and T. Popiel, *Quadratures and cubics in SO(3) and SO(1, 2)*, IMA J. Math. Control Inform., 23 (2006), pp. 463–474.

[32] L. Noakes and T. Popiel, *Geometry for robot path planning*, Robotica, 25 (2007), pp. 691–701.

[33] Y. Ohtake, A. Belyaev, M. Alexa, G. Turk, and H.-P. Seidel, *Multi-level partition of unity implicits*, ACM Trans. Graphics, 22 (2003), pp. 463–470.

[34] P. Perona, *Orientation diffusions*, IEEE Trans. Image Process., 7 (1998), pp. 457–467.

[35] I. U. Rahman, I. Drori, V. C. Stodden, D. L. Donoho, and P. Schröder, *Multiscale representations for manifold-valued data*, Multiscale Model. Simul., 4 (2005), pp. 1201–1232.

[36] B. Schölkopf and A. Smola, *Learning with Kernels*, MIT Press, Cambridge, MA, 2002.

[37] B. Schölkopf, F. Steinke, and V. Blanz, *Object correspondence as a machine learning problem*, in Proceedings of the 22nd International Conference on Machine Learning (ICML), Bonn, Germany, 2005, pp. 777–784.

[38] R. Shepard, *Multidimensional scaling, tree-fitting, and clustering*, Science, 210 (1980), pp. 390–398.

[39] F. Steinke and M. Hein, *Non-parametric regression between Riemannian manifolds*, in Advances in Neural Information Processing Systems 21, B. Schölkopf, J. Platt, and T. Hofmann, eds., MIT Press, Cambridge, MA, 2009, pp. 1561–1568.

[40] F. Steinke, M. Hein, J. Peters, and B. Schölkopf, *Manifold-valued thin-plate splines with applications in computer graphics*, Computer Graphics Forum, 27 (2008), pp. 437–448.

[41] F. Steinke, B. Schölkopf, and V. Blanz, *Support vector machines for* 3D *shape processing*, Computer Graphics Forum, 24 (2005), pp. 285–294.

[42] J. Tenenbaum, V. Silva, and J. Langford, *A global geometric framework for nonlinear dimensionality reduction*, Science, 290 (2000), pp. 2319–2323.

[43] Biometric Systems Lab (University of Bologna), Pattern Recognition and Image Processing Laboratory (Michigan State University), and Biometric Test Center (San Jose State University), *FVC* 2004: *The Third International Fingerprint Verification Competition*; Web site at http://bias.csr.unibo.it/fvc2004.

[44] H. Urakawa, *Calculus of Variations and Harmonic Maps*, AMS, Providence, RI, 1993.

[45] G. Wahba, *Spline Models for Observational Data*, CBMS-NSF Regional Conf. Ser. in Appl. Math. 59, SIAM, Philadelphia, 1990.

[46] R. Wald, *General Relativity*, The University of Chicago Press, Chicago, 1984.

[47] J. Wallner, H. Pottmann, and M. Hofer, *Fair webs*, The Visual Computer, 23 (2007), pp. 83–94.

[48] C. Wang, *Stationary biharmonic maps from* $\mathbb{R}^m$ *into a Riemannian manifold*, Comm. Pure Appl. Math., 57 (2004), pp. 419–444.

[49] H. Wendland, *Scattered Data Approximation*, Cambridge University Press, Cambridge, UK, 2005.

[50] R. Zayer, C. Rössl, and H.-P. Seidel, *Setting the boundary free: A composite approach to surface parameterization*, ACM Internat. Conf. Proc. Ser. 255, ACM, New York, 2005, pp. 91–100.