Image and Surface Processing

Behrend Heeren and Martin Rumpf

September 5, 2018

Contents

Preface These notes have been written to prepare the lecture course *Image and Surface Processing* which we taught at the Institute for Numerical Simulation, University of Bonn, in summer term 2018.

0 Introduction

Surface processing. The primal focus of the first part of this course is the mathematical foundation of *deformations and modeling tasks (animation)* of complex geometric objects represented as *discrete surfaces*, i.e. triangle meshes. In particular, we are interested in dynamics i.e. deformation paths and nontrivial motions. To this end, we assume that we are already given a nice triangulation (parametrization) without topological errors, geometrical noise or mesh artefacts. To enable efficient algorithms we build on state-of-the-art *simplification and approximation* techniques to design a *hierarchical/multiresolution method*.

In detail, we will investigate:

- local geometry: geometric quantities on a single (discrete) surface, in particular the notion of discrete curvature
- physics: nonlinear deformations based on a physically sound thin shell model and a corresponding discrete analgon (discrete shells)
- global geometry and dynamics: deformation paths and navigation in the space of discrete surfaces which is considered as a Riemannian manifold

1 Smooth and discrete surfaces

There are two major classes of surface representations: explicit (or parametric) and implicit surfaces. Parametric surfaces are defined by a vector-valued parametrization function

$$f : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^3 ,$$

that maps a two-dimensional domain $\Omega \subset \mathbb{R}^2$ to the surface

$$\mathcal{S} = f(\Omega) .$$

Implicit surfaces are defined as the zero-levelset of a scalar valued function

$$F : \mathbb{R}^3 \to \mathbb{R} ,$$

such that

$$\mathcal{S} = \{x \in \mathbb{R}^3 \,:\, F(x) = 0\} .$$

Example (Torus, $0 < r < R$)

$$f(\theta, \phi) = \begin{pmatrix} (R + r\cos\theta)\cos\phi \\ (R + r\cos\theta)\sin\phi \\ r\sin\theta \end{pmatrix} , \quad (\theta, \phi) \in \Omega = [0, 2\pi)^2 ,$$
$$F(x, y, z) = (x^2 + y^2 + z^2 + R^2 - r^2)^2 - 4R^2(x^2 + y^2) .$$

Note: For more complex shapes it is often necessary to split the domain into several patches (consistent/smooth transition has to be guaranteed, cf. definition of a manifold).

In this course we will focus on parametric surfaces (due to applications considered here!).

1.1 Differential geometry of parametric surfaces

In this chapter we gather basic properties of parametric surfaces $\mathcal{S} \subset \mathbb{R}^3$. For further reading we refer to [Bär00, dC76] (*cf*. also third chapter in [BKP+10]).

Definition 1.1. (Regular surface) The set $\mathcal{S} \subset \mathbb{R}^3$ is a *regular surface* if for each $p \in \mathcal{S}$ there is an $\epsilon > 0$, an open set $\Omega \subset \mathbb{R}^2$ and a smooth mapping $x : \Omega \to \mathbb{R}^3$, such that

(i) $x(\Omega) = \mathcal{S} \cap B_\epsilon(p)$ and $x : \Omega \to \mathcal{S} \cap B_\epsilon(p)$ is a homeomorphism.

(ii) The Jacobi matrix $D_\xi x \in \mathbb{R}^{3,2}$ has rank 2 for each $\xi \in \Omega$.

Remark: In the following we will assume that ϵ is large enough such that $\mathcal{S} \cap B_\epsilon(p) = \mathcal{S}$, *i.e.* there is a global *parametrization* $x : \Omega \to \mathcal{S}$.

Let us consider some $\xi = (\xi_1, \xi_2) \in \Omega$ and let $p = x(\xi) \in \mathcal{S}$. Then we define

$$T_p\mathcal{S} = \{\, \dot\gamma(0) \,|\, \gamma : (-1, 1) \to \mathcal{S},\ \gamma(0) = p\} .$$

Note that by the same definition we have

$$T_\xi\Omega = \{\dot\alpha(0) \,|\, \alpha : (-1, 1) \to \Omega,\ \alpha(0) = \xi\} = \mathbb{R}^2 .$$

If we consider $\alpha : (-1, 1) \to \Omega$ with $\alpha(0) = \xi$ and $\gamma_\alpha := x \circ \alpha$ we have $\dot\gamma_\alpha(0) = Dx\dot\alpha(0)$ and get

$$T_p\mathcal{S} = Dx\, T_\xi\Omega = \operatorname{span}\{\partial_{\xi_1} x(\xi), \partial_{\xi_2} x(\xi)\} .$$

In the following, we will denote the tangent vectors

$$V_i = \partial_i x(\xi) = \partial_{\xi_i} x(\xi) \,= \frac{\partial x(\xi)}{\partial \xi_i} = Dx e_i, \quad i = 1, 2 ,$$

as *canonical basis* of $T_p\mathcal{S}$, where $e_1, e_2 \in \mathbb{R}^2$ are the standard basis vectors of $\mathbb{R}^2$.

1.2 First and second fundamental forms

Definition 1.2 (First fundamental form)**.** The first fundamental form in $p \in \mathcal{S}$ is given by

$$g_p : T_p\mathcal{S} \times T_p\mathcal{S} \to \mathbb{R}, \quad g_p(U, V) := \langle U, V \rangle_{\mathbb{R}^3} .$$

After choosing a basis of $T_p\mathcal{S}$—here and in the following the canonical basis (V_1, V_2)—we can represent g_p by a symmetric, positive-definite matrix $g = g_\xi \in \mathbb{R}^{2,2}$ with

$$g_{ij} = g_p(V_i, V_j) = \langle V_i, V_j \rangle_{\mathbb{R}^3} , \tag{1.1}$$

i.e. we have $g = Dx^T Dx$. The pull-back of g_p to the parameter domain $\Omega \subset \mathbb{R}^2$ is defined as

$$g_\xi(u, v) = g_p(Dx\, u, Dx\, v) = u^T g v , \quad u, v \in T_\xi\Omega = \mathbb{R}^2 .$$

Geometrically, the first fundamental form is necessary to measure on the surface, *e.g.* to determine lengths of curves or angles between tangent vectors. Let $\gamma_\alpha = x \circ \alpha$ be as above, then the length of γ_α is defined as

$$\mathcal{L}[\gamma_\alpha] = \int_{-1}^{1} |\dot\gamma_\alpha(t)| \, \mathrm{d}t = \int_{-1}^{1} \sqrt{\langle Dx\dot\alpha(t), Dx\dot\alpha(t)\rangle_{\mathbb{R}^3}} \, \mathrm{d}t = \int_{-1}^{1} \sqrt{\langle Dx^T Dx\dot\alpha(t), \dot\alpha(t)\rangle_{\mathbb{R}^3}} \, \mathrm{d}t ,$$

where we actually have $Dx = Dx(\alpha(t))$.

To simplify notation, we will often drop the index and write $g = g_p$ or $g = g_\xi$, respectively. In particular, g refers to the bilinear form as well as to its representative matrix in $\mathbb{R}^{2,2}$. Note that $g \in \mathbb{R}^{2,2}$ is invertible, since $\mathcal{S} \subset \mathbb{R}^3$ is assumed to be regular.

For some $A \subset \Omega$ and for some function $\varphi : \mathcal{S} \to \mathbb{R}$ we have

$$\int_{x(A)} \varphi \, \mathrm{d}a = \int_A (\varphi \circ x)(\xi) \sqrt{\det g_\xi} \, \mathrm{d}\xi ,$$

and in particular for $\varphi \equiv 1$ we get

$$\mathrm{vol}(x(A)) = \int_A \sqrt{\det g_\xi} \, \mathrm{d}\xi .$$

Differentiation. For a function $\varphi : \mathcal{S} \to \mathbb{R}$ we define the differential $d_p\varphi$ as a linear form acting on tangent vectors $V \in T_p\mathcal{S}$ as directional derivative, i.e.

$$d_p\varphi(V) := \frac{d}{dt}\varphi(\gamma(t))\Big|_{t=0}$$

for an arbitrary curve $\gamma : (-1, 1) \to \mathcal{S}$ with $\gamma(0) = p$ and $\dot\gamma(0) = V$. For a vector-valued deformation $\phi : \mathcal{S} \to \mathbb{R}^3$ the definition above holds for each component of $\phi = (\phi_1, \phi_2, \phi_3)$. In particular, $d_p\phi$ defines a linear map between the tangent spaces, i.e.

$$d_p\phi : T_p\mathcal{S} \to T_{\phi(p)}\phi(\mathcal{S}) .$$

Definition 1.3 (Normal field)**.** Let $S^2 \subset \mathbb{R}^3$ be the 2-dimensional unit sphere. The (unit) normal field of $\mathcal{S}$ is a mapping $n : \mathcal{S} \to S^2$ with $n(p) \perp T_p\mathcal{S}$ for all $p \in \mathcal{S}$. We say that $\mathcal{S}$ is orientable if there is a continuous normal field. In particular, as $\mathrm{rank}(Dx) = 2$, we will write

$$n(p) = (n \circ x)(\xi) = \frac{x_{,1} \times x_{,2}}{|x_{,1} \times x_{,2}|}(\xi) .$$

Definition 1.4. (Shape operator) Let $\mathcal{S} \subset \mathbb{R}^3$ be regular and orientable, $p \in \mathcal{S}$. The shape operator $S_p : T_p\mathcal{S} \to T_p\mathcal{S}$ at p is the linear mapping defined via $S_p(U) = d_p n(U)$ for $U \in T_p\mathcal{S}$.

Remark: As $T_{n(p)}S^2 = n(p)^\perp = T_p\mathcal{S}$ the shape operator S_p is indeed an endomorphism on $T_p\mathcal{S}$.

Definition 1.5. (Second fundamental form) Let $\mathcal{S} \subset \mathbb{R}^3$ be regular and orientable, $p \in \mathcal{S}$. The second fundamental form $h = h_p$ is the bilinear form on $T_p\mathcal{S}$ associated with S_p, *i.e.*

$$h_p(U, V) := g_p(S_p U, V)\,, \quad U, V \in T_p\mathcal{S}\,.$$

The corresponding matrix representation $h = h_\xi \in \mathbb{R}^{2,2}$ is given by

$$h_{ij} = h_p(V_i, V_j) = g_p(S_p V_i, V_j) = \langle d_p n\, \partial_i x, \partial_j x\rangle_{\mathbb{R}^3} = \langle \partial_i(n \circ x), \partial_j x\rangle_{\mathbb{R}^3}\,. \tag{1.2}$$

Note that since $n(p) \perp T_p\mathcal{S}$ we have

$$0 = \partial_{\xi_i}\Big(g_p(n \circ x, \partial_j x)\Big) = g_p\left(\partial_i(n \circ x), \partial_j x\right) + g_p\left((n \circ x), \partial_i\partial_j x\right) = h_p(V_i, V_j) + g_p\left((n \circ x), \partial^2_{ij} x\right)\,,$$

hence $h \in \mathbb{R}^{2,2}$ is symmetric and we have

$$h_{ij} = -\langle n \circ x, \partial^2_{ij} x\rangle_{\mathbb{R}^3}\,.$$

Finally, we can represent the (symmetric) matrix $h \in \mathbb{R}^{2,2}$ by

$$h = Dn^T Dx\,, \quad Dn = \left[\frac{\partial n(p)}{\partial \xi_1}, \frac{\partial n(p)}{\partial \xi_2}\right] \in \mathbb{R}^{3,2}\,.$$

If we write S_p in the canonical basis (V_1, V_2), *i.e.*

$$S_p V_i = \sum_{k=1}^{2} s_{ki} V_k$$

for $i = 1, 2$, the coefficient matrix $s = s_\xi \in \mathbb{R}^{2,2}$ is the representation of S_p in the parameter domain. Since

$$h_{ij} = g_p(S_p V_i,\, V_j) = g_p\Big(\sum_{k=1}^{2} s_{ki} V_k,\; V_j\Big) = \sum_{k=1}^{2} s_{ki}\; g_p\Big(V_k, V_j\Big) = \sum_{k=1}^{2} s_{ki} g_{kj},$$

we get $h = s^T g$, i.e. due to the symmetry of g and h we have

$$s_\xi = g_\xi^{-1} h_\xi\,. \tag{1.3}$$

Remark on notation: S_p denotes either the endomorphism on $T_p\mathcal{S}$ or the corresponding matrix $S_p \in \mathbb{R}^{3,3}$, whereas $s \in \mathbb{R}^{2,2}$ denotes the matrix representation of S_p in the canonical basis.

Since g and h are symmetric forms on $T_p\mathcal{S}$ we get for $U, V \in T_p\mathcal{S}$

$$g_p(S_p U, V) = h_p(U, V) = h_p(V, U) = g_p(S_p V, U) = g_p(U, S_p V)\,,$$

which means that S_p is symmetric with respect to the metric. Hence S_p and thus s_ξ diagonalize in an orthonormal basis.

Definition 1.6 (Curvatures)**.** The eigenvalues κ_1, κ_2 of s_ξ are denoted as principal curvatures of $\mathcal{S}$ in $p = x(\xi)$. The mean curvature in p is defined as the sum $H_p = \operatorname{tr} s_\xi = \kappa_1 + \kappa_2$ and the Gaussian curvature in p as the product $K_p = \det s_\xi = \kappa_1 \cdot \kappa_2$.

Note that $\det S_p = 0$ since there is no normal variation in normal direction, hence the eigenvalues of S_p are given by $0, \kappa_1, \kappa_2$.

The *normal curvature* of $p \in \mathcal{S}$ in some direction $U \in T_p\mathcal{S}$ is defined as

$$\kappa_p(U) = \frac{h_p(U, U)}{g_p(U, U)}\,.$$

Intuitively, $\kappa_p(U)$ describes the curvature of a curve $\gamma : I \subset (-1, 1) \to \mathcal{S}$ with $\gamma(0) = p$ and $\gamma(I) = \mathcal{S} \cap (T_p\mathcal{S})^\perp$ at $t = 0$. If one obeys the ordering convention $\kappa_1 \le \kappa_2$, one can show that

$$\kappa_1 = \min_{U \in T_p\mathcal{S}} \kappa_p(U)\,, \quad \kappa_2 = \max_{U \in T_p\mathcal{S}} \kappa_p(U)\,.$$

1.3 The relative shape operator

Later, we aim at measuring differences between two (discrete) surfaces up to rigid body motions. That means, if $\mathcal{S} \subset \mathbb{R}^3$ is a parametric surface and $\phi : \mathcal{S} \to \mathbb{R}^3$ is a deformation, we aim at quantifying the *dissimilarity* of $\mathcal{S}$ and $\tilde{\mathcal{S}} := \phi(\mathcal{S})$ up to rigid body motions.

Theorem 1.7 (The Fundamental Theorem of Surfaces)**.** *Congruent parametric surfaces in $\mathbb{R}^3$ have the same first and second fundamental forms. Conversely, two parametric surfaces in $\mathbb{R}^3$ with the same first and second fundamental forms are congruent.*

Hence the theorem above suggests to measure differences of first fundamental forms g and $\tilde{g}$ as well as second fundamental forms h and $\tilde{h}$ of $\mathcal{S}$ and $\tilde{\mathcal{S}}$, respectively. First, it is easy to see that $g = \tilde{g}$ if $D\phi^T D\phi = \mathbb{1}$. Indeed, if $x : \Omega \to \mathbb{R}^3$ is a parametrization of $\mathcal{S}$ then $\tilde{x} = \phi \circ x$ is a parametrization of $\tilde{\mathcal{S}}$, and hence $\tilde{g} = D\tilde{x}^T D\tilde{x} = Dx^T D\phi^T D\phi Dx$. Second, since the shape operator represents the second fundamental form in the metric, one often penalizes deviations in the shape operator. Furthermore, if $g = \tilde{g}$ then $h = \tilde{h}$ iff. $s = \tilde{s}$, cf. (1.3). In the most general setup one aims at comparing the embedded shape operators $S_p : T_p\mathcal{S} \to T_p\mathcal{S}$ and $\tilde{S}_{\tilde{p}} : T_{\tilde{p}}\tilde{\mathcal{S}} \to T_{\tilde{p}}\tilde{\mathcal{S}}$, for an arbitrary point $p \in \mathcal{S}$ and $\tilde{p} = \phi(p)$. However, since these operators live on different tangent spaces one defines:

Definition 1.8 (Pulled-back shape operator)**.** The pulled-back shape operator $S_p^*[\phi] : T_p\mathcal{S} \to T_p\mathcal{S}$ is given by

$$g_p\left(S_p^*[\phi]U,\, V\right) = h_{\phi(p)}\left(D\phi U,\, D\phi V\right)\,, \quad \forall U, V \in T_p\mathcal{S}\,. \tag{1.4}$$

Definition 1.9 (Relative shape operator)**.** The *relative shape operator* $S_p^{\text{rel}}[\phi]$ is defined as the pointwise difference, *i.e.*

$$S_p^{\text{rel}}[\phi] : T_p\mathcal{S} \to T_p\mathcal{S}\,, \quad S_p^{\text{rel}}[\phi] := S_p - S_p^*[\phi]\,. \tag{1.5}$$

The matrix representations $s_\xi^*[\phi] \in \mathbb{R}^{2,2}$ and $s_\xi^{\text{rel}}[\phi] \in \mathbb{R}^{2,2}$ of $S_p^*[\phi]$ and $S_p^{\text{rel}}[\phi]$, respectively, are given by

$$s_\xi^*[\phi] = g_\xi^{-1}\tilde{h}_\xi\,, \quad s_\xi^{\text{rel}}[\phi] = s_\xi - s_\xi^*[\phi] = g_\xi^{-1}(h_\xi - \tilde{h}_\xi)\,. \tag{1.6}$$

Remark: The fundamental theorem of surfaces provides a *geometric* argument why to measure differences in first and second fundamental forms. Later, we will also consider a *physical* justification.

1.4 Intrinsic vs. extrinsic surface quantities

Definition 1.10. (Isometric surfaces) A differentiable mapping $\phi : \mathcal{S} \to \tilde{\mathcal{S}}$ between two surfaces $\mathcal{S} \subset \mathbb{R}^3$ and $\tilde{\mathcal{S}} \subset \mathbb{R}^3$ is a local isometry, if for each $p \in \mathcal{S}$ the differential $d_p\phi : T_p\mathcal{S} \to T_{\phi(p)}\tilde{\mathcal{S}}$ is a linear isometry with respect to the first fundamental forms, i.e.

$$g_p(V, W) = \tilde{g}_{\phi(p)}\Big(d_p\phi(V),\, d_p\phi(W)\Big) \quad \forall V, V \in T_p\mathcal{S}\,.$$

If a local isometry $\phi : \mathcal{S} \to \tilde{\mathcal{S}}$ is bijective we say that ϕ is an *isometry*. If there is an isometry $\phi : \mathcal{S} \to \tilde{\mathcal{S}}$, then the two surfaces $\mathcal{S}, \tilde{\mathcal{S}}$ are said to be *isometric*.

A quantity of a surface, *e.g.* some function $f_{\mathcal{S}} : \mathcal{S} \to \mathbb{R}$, is said to be *intrinsic*, if it is not distorted under local isometries, *i.e.* we have $f_{\mathcal{S}} = f_{\tilde{\mathcal{S}}} \circ \phi$ for every local isometry $\phi : \mathcal{S} \to \tilde{\mathcal{S}}$. Otherwise, a surface quantity is said to be *extrinsic*.

Remark: An intrinsic quantity only depends on the first fundamental form.

Obviously, the first fundamental form is intrinsic, as well as the Laplace-Beltrami operator. However, mean curvature (and thus also the principle curvatures) is an *extrinsic* quantity, as, for example, $H \equiv 0$ for the plane but $H \equiv 1$ for the cylinder, whereas plane and cylinder are locally isometric. Likewise, the second fundamental form as well as the shape operator are also extrinsic quantities. However, the Gaussian curvature is *not* extrinsic:

Theorem 1.11 (Theorema egregium (Gauss))**.** *The Gaussian curvature is an intrinsic invariant of a surface.*

This implies that the sphere cannot be unfolded onto a flat plane without distorting the distances, i.e. a sphere and a plane are not isometric, even locally. This fact is of enormous significance for cartography: it implies that no planar (flat) map of Earth can be perfect, even for a portion of the Earth's surface.

Later, we want to quantify distortions induced by isometric deformations. The results above imply that this can be done by measuring differences in shape operators (in a suitable matrix norm) or, even simpler, differences in mean curvature. On the other hand, measuring differences in Gaussian curvature does not make sense due to Gauss' theorem. Hence a particular focus will be on deriving a discrete notion of mean curvature whereas the discrete notion of Gaussian curvature is less important.

Nevertheless, an important property of Gaussian curvature of an surface is the relation to its topology as stated in the Gauss-Bonnet theorem:

Theorem 1.12 (Theorem of Gauss-Bonnet)**.** *For a compact, orientable, closed surface $\mathcal{S} \subset \mathbb{R}^3$ we have*

$$\int_{\mathcal{S}} K \, \mathrm{d}a = 2\pi \cdot \chi_{\mathcal{S}} \, ,$$

where the Euler-characteristic $\chi_{\mathcal{S}} \in \mathbb{N}$ is given as $\chi_{\mathcal{S}} = 2(1-g)$, where g denotes the genus of the surface here.

1.5 Unstructured triangle meshes

A natural choice for parametrizing functions are polynomials, because they can be evaluated by elementary arithmetic operations. Furthermore, due to the Weierstrass Theorem, each smooth function can be approximated by a polynomial up to any desired precision (on a compact domain). From Taylor's theorem we know that a smooth function on a compact interval of length h can be approximated by a polynomial of degree p such that the approximation error behaves like $O(h^{p+1})$. As a consequence, there are two possibilities to improve the accurancy of the approximation: one can either raise the degree of the polynomial (p-refinement) or partition the domain in small patches, i.e. use more segments for the approximation (h-refinement).
In *Geometry Processing* applications one usually prefers h-refinement. With the today's computer architecture, processing a large number of of simple objects is often easier and more efficient that processing a smaller number of more complex objects. This leads to the somewhat extremal choice of C^0 piecewise linear surface representations, i.e. polygonal meshes, which have become the widely established standard in geometry processing.
Hence we consider $p = 1$ and $h \to 0$, i.e. we are dealing with an approximation power of $O(h^2)$ which is controlled by second-order information of the surface, i.e. curvature bounds.

Unstructured triangle meshes allow for arbitrary connectivities. Furthermore, geometric details, feature creases etc can be represented easily, hence unstructured meshes provide a higher flexibility and still allow for efficient surface processing. A triangle mesh $\mathcal{M}_h$ consists of geometric and topological components, whereas the latter one can be represented by a graph structure, i.e. we have a set of vertices

$$\mathcal{V} = \{v_1, \ldots, v_n\}$$

and a set of triangular faces

$$\mathcal{F} = \{f_1, \ldots, f_m\} \subset \mathcal{V} \times \mathcal{V} \times \mathcal{V} \, .$$

Based on $\mathcal{V}$ and $\mathcal{F}$ one can deduce a set of edges

$$\mathcal{E} \subset \mathcal{V} \times \mathcal{V} \, .$$

The geometric embedding of a triangle mesh is given by a mapping

$$E : \mathcal{V} \to \mathbb{R}^3 \, , \quad E(v_i) = X_i \, .$$

An important topological quality of the mesh is whether or not it is a (discrete) 2-manifold. This is the case if it contains neither non-manifold edges nor non-manifold vertices nor self-intersections.

The famous *Euler formula* states an interesting relation between the number of vertice, faces and edges in a closed and connected (but otherwise unstructured) mesh:

$$\chi_{\mathcal{M}_h} := |\mathcal{V}| + |\mathcal{F}| - |\mathcal{E}| = 2(1-g)\,, \tag{1.7}$$

where g is here the genus of the surface (intuitively, the genus counts the number of handles of a surface). Since for most applications g is small, one can neglect the left hand side. Furthermore, as each edge is incident to two faces and each face has three edges, we have $3|\mathcal{F}| = 2|\mathcal{E}|$ and one can deduce the following mesh statistics:

- the number of triangles is twice the number of vertices, i.e. $|\mathcal{F}| \approx 2\,|\mathcal{V}|$,
- the number of edges is three times the number of vertices, i.e. $|\mathcal{E}| \approx 3\,|\mathcal{V}|$,
- the average vertex valence (number of incident edges) is 6.

As mentioned above, a triangular mesh is uniquely determined by its *geometry* and its *connectivity*. The geometry is described by the embedding $E : \mathcal{V} \to \mathbb{R}^3$, *i.e.* the coordinates of the vertices, and the connectivity is encoded in the set of faces $\mathcal{F} \subset \mathcal{V} \times \mathcal{V} \times \mathcal{V}$ or equivalently in the set of edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. Note that all further structural properties of the mesh, such as neighbouring relationships, boundary etc., can be derived from $\mathcal{F}$ or $\mathcal{E}$, respectively.

Geometry. If $n \in \mathbb{N}$ denotes the number of vertices in the mesh, we often identify a vertex $X_i = E(v_i)$ with $v_i \in \mathcal{V}$ or with its global index $i \in \{1, \dots, n\}$. Likewise, if $m \in \mathbb{N}$ denotes the number of faces in the mesh, we often identify a face $f_j \in \mathcal{F}$ with its global index $j \in \{1, \dots, m\}$. Moreover, we might represent a face $f = (i_0, i_1, i_2)$ by its embedded triangle $T = T(f)$, i.e.

$$T = T(f) = (E(v_{i_0}),\, E(v_{i_1}),\, E(v_{i_2})) = (X_{i_0},\, X_{i_1},\, X_{i_2}) \quad \subset \mathbb{R}^3\,. \tag{1.8}$$

Besides its global index $i \in \{1, \dots, n\}$, each vertex $X = X_i$ belongs to at least one face f, hence it has an additional local index $j \in \{0, 1, 2\}$ with respect to f. That means, we sometimes rewrite (1.8) as

$$T = T(f) = \Big(X_0(f), X_1(f), X_2(f)\Big) \quad \subset \mathbb{R}^3\,. \tag{1.9}$$

Analogously, each edge E of a mesh belongs to at least one face and hence it also has a local index $j \in \{0, 1, 2\}$ with respect to f, *i.e.* $E = E_j(f)$. Moreover, we make use of the convention that an edge with local index j (wrt. face f) connects the nodes with local indices $j-1$ and $j+1$ (wrt. face f), where the notation is modulo 3, *i.e.*

$$E_j = X_{j-1} - X_{j+1}\,, \quad j \in \{0, 1, 2\} \bmod 3\,.$$

That means E_j is opposite X_j in face f, if j is the local index wrt. face f. In the following, it will be clear from the context if we are referring to the global or local index of a vertex or edge, respectively.

Topology. We assume that each mesh is a *two-dimensional discrete manifold*, or (discrete) 2-manifold, in the sense of [DKT08]. This is the case if it contains neither non-manifold edges nor non-manifold vertices nor self-intersections. That means for example, that any pair of triangles either shares one edge or one node or that their intersection is empty. In particular, we do not allow for hanging nodes, *i.e.* nodes that do not belong to any face, or degenerated faces, *i.e.* faces with less than three different nodes.

Orientation. The order of the local indices of nodes within one face determines the orientation of the face and hence of the mesh. Since we only consider (approximations of) orientable surfaces we assume that all local indices are ordered consistently.

Definition 1.13 (Discrete surface)**.** A *discrete surface* is a triangular mesh $\mathcal{M}_h = (\mathcal{V}, \mathcal{F}, \mathcal{E})$ along with an injective embedding $E : \mathcal{V} \to \mathbb{R}^3$, such that $\mathcal{M}_h$ is a discrete 2-manifold which is orientable (*i.e.* has consistent local index ordering).

Remark: A discrete surface is entirely described by the pairing $\mathcal{M}_h = (E(\mathcal{V}), \mathcal{F})$, where $E(\mathcal{V}) \in \mathbb{R}^{3n}$.

Parametrization. We assume that each discrete surface $\mathcal{M}_h$ is parametrized over a reference or parameter domain Ω_h. Yet different from the continuous setting this reference domain is not a connected subset of $\mathbb{R}^2$ but rather an abstract collection of multiple reference triangles as it is often used in the context of subdivision surfaces (*cf.* [Rei95]). Thus each face $f \in \mathcal{F}$ of $\mathcal{M}_h$, with the corresponding embedded/geometric triangle $T = T(f)$ as in (1.8) resp. (1.9) is parametrized over a reference triangle given by the *unit triangle*

$$\omega := (\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}) \subset \mathbb{R}^2$$

via an affine mapping $X_f : \omega \to T$, which is defined by

$$X_f(\xi) := X_f(\xi_1, \xi_2) := \xi_1 X_1(f) + \xi_2 X_2(f) + (1 - \xi_1 - \xi_2) X_0(f) \tag{1.10}$$

for the barycentric coordinates $\xi \in \omega$, *i.e.* $\xi = (\xi_1, \xi_2)$ with $0 \leq \xi_1, \xi_2 \leq 1$ and $\xi_1 + \xi_2 \leq 1$. Formally, the reference domain is given by $\Omega_h = \omega \times \mathcal{F}$, a global parametrization via $X : (\xi, i) \mapsto X_{f_i}(\xi)$. Wherever it is possible, we drop the dependence of the local parametrization X on the face f_j in the following and write $X = X_{f_j}$.

Discrete first fundamental form. Let $\mathcal{S} \subset \mathbb{R}^3$ be a regular embedded surface with (local) parametrization $x : \Omega \to \mathcal{S}$ and $\xi \in \Omega$. Following Def. 1.2 resp. (1.1) we can represent the first fundamental form g at some point $x(\xi)$ by the matrix $g_\xi = Dx(\xi)^T Dx(\xi)$. According to (1.10), the local parametrization X of a a discrete surface $\mathcal{M}_h$ is affine, hence its derivative is constant on each triangle $f \in \mathcal{M}_h$, *i.e.*

$$DX|_f = \left(\frac{\partial X_f}{\partial \xi_1}, \frac{\partial X_f}{\partial \xi_2} \right) = \Big[X_1(f) - X_0(f) \,\Big|\, X_2(f) - X_0(f) \Big] = \Big[E_2(f) \,\Big|\, -E_1(f) \Big] \in \mathbb{R}^{3,2} . \tag{1.11}$$

Hence the definition of an elementwise constant discrete first fundamental form follows canonically:

$$G_f = (DX|_f)^T DX|_f \in \mathbb{R}^{2,2} , \quad f \in \mathcal{F} . \tag{1.12}$$

To simplify notation we will often drop the dependence on f and write $G = G_f$. Note that $\det G_f = 0$ iff. f has parallel edges, which is not admissible due to the assumption that $\mathcal{M}_h$ is a discrete 2-manifold. Hence G_f is invertible for each $f \in \mathcal{F}$. Furthermore, we get the following formula for the area $a_f = |T|$ of an embedded triangle:

$$a_f = \int_T \mathrm{d}a = \int_\omega \sqrt{\det G_f}\, \mathrm{d}\xi = \frac{1}{2} \sqrt{\det G_f} . \tag{1.13}$$

1.6 Discrete differential geometry

Differential and geometric quantities defined on a surface require the surface to be sufficiently smooth, e.g. the definition of curvatures requires the existence of second derivatives. Since polygonal meshes are piecewise affine and globally only of class C^0, many of the concepts presented in Sec. 1.1 can not be applied directly. However, usually a discrete surface (i.e. a polygonal mesh) is assumed to be an approximation of a smooth surface. Hence one aims to compute approximations of differential and geometric properties of the smooth surface directly from the mesh data. This leads to the derivation of discrete equivalents of the geometric notions of classical differential geometry. In particular, the study of the discrete equivalents themselves defines a new and active mathematical field, namely *Discrete Differetial Geometry (DDG)*. The guiding principles of DDG:

- **Weak/integrated notions** Higher order quantities, such as curvatures, are defined in an integrated or weak sense. In particular, on general meshes convergence is mostly shown in a weak or integrated sense.
- **Spatial averaging** To get a pointwise evaluation, e.g. at a vertex, one computes the integrated quantity in a neighbourhood of that vertex and divides by the associated area (cf. [MDSB02])
- **Attach quantities to appropriate locations**, i.e. associate physical/geometric quantities at their natural locations, not necessarily at vertices. Here the field of Discrete Exterior Calculus (DEC) is an appropiate tool (not considered here, cf. [DKT08, Hir03, DHLM05]).
- **Discretize theory, not equations!** That means, one aims at a consistent discrete theory. To ensure the existence of fundamental properties (e.g. Gauss-Bonnet), one often uses *top-down* instead of *bottom-up* approaches, e.g. one directly defines a notion of discrete curvature without having a notion of a discrete normal field in the first place.

In later applications we primarily want to make use of extrinsic curvature measures, i.e. we focus here mainly on a discrete notion of mean curvature and the shape operator.
Having in mind Gauss' Theorema Egregrium, a simple consideration of two triangles glued together at one edge reveals that on a discrete surface:

- mean curvature is concentrated at edges,
- Gaussian curvature is concentrated at vertices.

Hence, a shape operator is also naturally associated with edges. Nevertheless, discrete shape operators are often associated with a triangle by taking into account the bending accross the three edges. Finally, one can also define shape operators on vertices e.g. by averaging over adjacent edges. In the following, after having introduced several concepts of normals on a mesh, we consider vertex-based and edge-based curvature quantities, where the primal focus is on the latter one.

Normals First, we define the weighted face normal $\tilde{N}_f \in \mathbb{R}^3$ resp. the unit face normal $N_f \in S^2$ on a triangle $T(f) = (X_0, X_1, X_2)$ by

$$\tilde{N}(f) = \tilde{N}_f = (X_1 - X_0) \times (X_2 - X_0)\,, \quad N(f) = N_f = \frac{(X_1 - X_0) \times (X_2 - X_0)}{\|(X_1 - X_0) \times (X_2 - X_0)\|}\,.$$

Since extrinsic bending is naturally measured across edges, a normal field associated with edges is a canonical choice. Usually, these edge normals are located at edge midpoints and computed as weighted average of the adjacent triangle normals, i.e. if f_l and f_r are the adjacent faces of some edge E we set

$$N_E = \frac{\alpha_l N_{f_l} + \alpha_r N_{f_r}}{\|\alpha_l N_{f_l} + \alpha_r N_{f_r}\|}\,, \tag{1.14}$$

where the weights might be chosen e.g. as $\alpha_r = \alpha_l = 1/2$. Another ansatz for defining an edge normal is given by postulating $N_E \perp E$ and then parametrizing N_E via the angle e.g. between N_E and $N(f_s)$, where $s \in \{l, r\}$. Computing vertex normals as spatial averages of face normals (alternatively, of edge normals) in a local 1-ring neighbourhood leads to normalized weighted average of the (constant) face normals, i.e.

$$N(v) = N_v = \frac{\sum_{f \in N_1(v)} \alpha_f N_f}{\| \sum_{f \in N_1(v)} \alpha_f N_f \|}\,.$$

There are numerous alternatives for the weights α_f, e.g. $\alpha_f = 1$ or $\alpha_f = a_f$ or $\alpha_f = \gamma_f$, where γ_f is the interior triangle angle in f at vertex v. For most applications, the latter angle-weighted vertex normal provides a good trade-off between computational efficiency and accurancy.

Discrete Gauss Curvature As mentioned above, discrete Gaussian curvature is supposed to be concentrated at vertices. Following the principles of DDG, one defines the discrete Gaussian curvature such that a discrete version of Gauss-Bonnet holds. That means for some area A_v associated with some vertex $v \in \mathcal{V}$ one defines an *integrated* Gaussian curvature by the so-called *angle-defect*, i.e.

$$\int_{A_v} K_h(x)\,\mathrm{d}a := 2\pi - \sum_{f : v \in f} \gamma_f\,, \tag{1.15}$$

where γ_f denotes the interior triangle angle in f at vertex v, and verifies immediately

$$\int_{\mathcal{M}_h} K_h(x)\,\mathrm{d}a = \sum_{v \in \mathcal{V}} \int_{A_v} K_h(x)\,\mathrm{d}a = 2\pi|\mathcal{V}| - \sum_{v \in \mathcal{V}} \sum_{f : v \in f} \gamma_f = 2\pi\left(|\mathcal{V}| - \frac{1}{2}|\mathcal{F}|\right) = 2\pi\left(|\mathcal{V}| + |\mathcal{F}| - |\mathcal{E}|\right),$$

where we have used $\sum_{v : v \in f} \gamma_f = \pi$ as well as $3|\mathcal{F}| = 2|\mathcal{E}|$ in the last step. Cohen-Steiner and Morvan [CSM03] prove the convergence of (1.15) to its continuous counterpart (in an integrated sense!). Following the paradigm of *spatial averaging* one might define Gaussian curvature evaluated at a vertex, i.e.

$$K_h(v) = \frac{\int_{A_v} K_h(x)\,\mathrm{d}a}{|A_v|} = \frac{1}{|A_v|}\Big(2\pi - \sum_{f : v \in f} \gamma_f\Big)\,.$$

Edge-based curvature measures As mentioned above a discrete notion of mean curvature is naturally associated with edges, more precisely, with the bending between two adjacent faces. Intuitively, the isometric bending of two adjacent triangles sharing an edge E is captured by the so-called *dihedral angle* θ_E, i.e.

$$\theta_E := \angle\Big(N(f_l) \times E,\, E \times N(f_r)\Big)\,, \quad E = T(f_r) \cap T(f_l)\,.$$

Hence we have $\theta_E = \pi$ if $N(f_l) \parallel N(f_r)$. Since bending in either direction should be penalized equally (in absolut value) with positive values if $\theta_E > \pi$ and negative if $\theta_E < \pi$, one intuitively expects

$$H_E \sim -2\,\cos\frac{\theta_E}{2}\,. \tag{1.16}$$

Note that sometimes the quantity $\theta_E - \pi$ is refered to as dihedral angle as well, then (1.16) is usually replaced by $H_E \sim 2\,\sin\frac{\theta_E}{2}$. A Taylor expansion of (1.16) about π leads to

$$H_E \sim (\theta_E - \pi) + O(|\theta_E - \pi|^3)\,.$$

Finally, based on H_E one can define an edge-based shape operator $S_E \in \mathbb{R}^{3,3}$ by

$$S_E = \frac{1}{3}\,H_E\,(N_E \times E) \otimes (N_E \times E)\,,$$

which satisfies $S_E E = 0$ (no curvature along the edge), $S_E N_E = 0$ (no bending in normal direction) and $\operatorname{tr} S_E = H_E$.

The discrete notion of mean curvature as stated in (1.16) can be found in several approaches, *cf*. *e.g*. [CSM03, BMF03, GHDS03, HP04, War06, Sul08], and will be justified in the following. Note, however, that in most approaches a discrete notion of mean curvature is first derived in an integrated sense, i.e. if d_E denotes an area associated with edge E satisfying $E \subset d_E \subset T(f_r) \cup T(f_l)$, one sets

$$\int_{d_E} H_E(x)\,\mathrm{d}a := -2\,\cos\frac{\theta_E}{2}\,\|E\|\,.$$

Convergence results by Cohen-Steiner and Morvan [CSM03] Let $V \subset \mathbb{R}^3$ a convex body such that $\mathcal{S} := \partial V$ is a smooth and compact surface. For $\epsilon > 0$ and some area $B \subset \mathcal{S}$ consider the offset

$$V_\epsilon(B) = \{x + \epsilon\,t\,n(x) : x \in B,\, t \in [0,1]\}\,.$$

Then *Steiner's tube formula* reads (*cf*. [CSM03, Sul08])

$$|V_\epsilon(B)| = \epsilon \int_B \mathrm{d}a + \frac{1}{2}\epsilon^2 \int_B H\,\mathrm{d}a + \frac{1}{3}\epsilon^3 \int_B K\,\mathrm{d}a\,, \tag{1.17}$$

which can be re-formulated as

$$\int_B H\,\mathrm{d}a = 2\epsilon^{-2}\Big(|V_\epsilon(B)| - \epsilon\,|B|\Big) + O(\epsilon)\,. \tag{1.18}$$

Now let $B_h \subset \mathcal{M}_h$. For illustrative reasons, we first consider $B_h = T_1 \cup T_2$. If $E = T_1 \cap T_2$, then we postulate $\theta_E > \pi$ since B_h is supposed to be (part of) the boundary of a *convex* body. For some $\epsilon > 0$ we define the offset $V_\epsilon(B_h)$ in a canonical way. If β_E denotes the angle between $N(f_1)$ and $N(f_2)$ we get

$$|V_\epsilon(B_h)| - \epsilon\,|B_h| = \frac{\beta_E}{2\pi} \cdot \epsilon^2\pi\,\cdot \|E\|,$$

which represents a $\beta_E/(2\pi)$-fraction of the volume of a cylinder of height $\|E\|$ and radius ϵ. Using (1.18) with $\beta_E = \theta_E - \pi$ suggests the definition

$$\int_{B_h} H_h\,\mathrm{d}a := (\theta_E - \pi)\,\|E\|\,.$$

For arbitrary $B_h \subset T_1 \cup T_2$ one obtains analogously

$$\int_{B_h} H_h \, \mathrm{d}a := (\theta_E - \pi) \, \|E \cap B_h\| \, ,$$

and for arbitrary $B_h \subset \mathcal{M}_h$ one obtains

$$\int_{B_h} H_h \, \mathrm{d}a = \sum_{E \subset B_h} (\theta_E - \pi) \, \|E \cap B_h\| \, + O(\epsilon),$$

since the cone-like volume $V_\epsilon(B_h)$ sitting at a vertex is given by a fraction of the ball of radius ϵ, hence scales as ϵ^3.

Definition [Delaunay triangulation] A Delaunay triangulation for a set of points in $\mathbb{R}^2$ is a triangulation such that no point is inside the circumcircle of any triangle.

Remark: The Delaunay triangulation of a discrete point set P in general position corresponds to the dual graph of the Voronoi diagram for P. The Voronoi region V_p associated with one point $p \in P$ is defined as

$$V_p = \{q \in \mathbb{R}^2 \, | \, \mathrm{dist}(p, q) \leq \mathrm{dist}(p', q) \quad \forall p' \in P\} \, .$$

The Voronoi diagram is simply the tuple of cells V_p for all $p \in P$.

Definition [ϵ-sample] A point set $P \subset \mathbb{R}^3$ is an ϵ-sample of a regular surface $\mathcal{S} \subset \mathbb{R}^3$ if for all $p \in \mathcal{S}$

$$B_{r(\epsilon,p)}(p) \cap P \neq \varnothing \, ,$$

where $r(\epsilon, p) = \epsilon \cdot \mathrm{dist}(p, \mathrm{med}(\mathcal{S}))$, where $\mathrm{med}(\mathcal{S})$ denote the medial axis of $\mathcal{S}$.

Cohen-Steiner and Morvan [CSM03] prove the following convergence result:

Theorem [Convergence of weak edge-based mean curvature [CSM03]] Let $\mathcal{S} \subset \mathbb{R}^3$ be a smooth surface, not necessarily the boundary of a convex body, $\mathcal{M}_h = (\mathcal{V}, \mathcal{F}, \mathcal{E})$ an approximating triangle mesh and $\epsilon > 0$ sufficiently small. If the set $\{E(v) | v \in \mathcal{V}\}$ is an ϵ-sample of $\mathcal{S}$ and the triangulation of $\mathcal{M}_h$ is Delaunay, then for any $B_h \subset \mathcal{M}_h$

$$\left| \sum_{E \subset B_h} (\theta_E - \pi) \, \|E \cap B_h\| \; - \int_{\Phi^{-1}(B_h)} H \, \mathrm{d}a \right| = O(\epsilon) \, , \tag{1.19}$$

where $\Phi : \mathcal{S} \to \mathcal{M}_h$ is the *shortest distance map* which is a homeomorphism if h is sufficiently small.

Remark: The proof is based on differential 2-forms and geometric measure theory and is beyond the scope of this course. A similar convergence result is shown for the discrete Gaussian curvature as defined in (1.15), as well as some discrete seecond fundamental form.

Triangle-averaged discrete shape operator We have already derived a matrix representation $G \in \mathbb{R}^{2,2}$ of a discrete first fundamental form that lives in the reference domain. Furthermore, $G = G_f$ is constant on each triangle f. Hence we aim at defining a matrix representation $H = H_f \in \mathbb{R}^{2,2}$ of a discrete second fundamental form that also lives in the reference domain and is elementwise constant. If we make use of (1.3) we can eventually derive a matrix representation $S = S_f \in \mathbb{R}^{2,2}$ of a discrete shape operator by setting $S_f = G_f^{-1} H_f$.

Let $f \in \mathcal{F}$ be an arbitrary triangle of a discrete surface $\mathbf{S}$ with a local parametrization $X = X_f$ as defined in (1.10). Note that we have

$$\frac{\partial X}{\partial \xi_1} = X_1 - X_0 = E_2 \, , \quad \frac{\partial X}{\partial \xi_2} = X_2 - X_0 = -E_1 \, . \tag{1.20}$$

Plugging this into (1.2) yields for the entries of $H = H_f$:

$$\begin{aligned} H_{11} &= \Big\langle dN(E_2), E_2 \Big\rangle_{\mathbb{R}^3} \quad H_{12} = -\Big\langle dN(E_2), E_1 \Big\rangle_{\mathbb{R}^3} \\ H_{21} &= -\Big\langle dN(E_1), E_2 \Big\rangle_{\mathbb{R}^3} \quad H_{22} = \Big\langle dN(E_1), E_1 \Big\rangle_{\mathbb{R}^3} \end{aligned}$$

For an edge $E \in \mathcal{M}_h$ we define an edge normal N_E by (1.14) using uniform weights $\alpha_r = \alpha_l = 1/2$. With normals associated to edge midpoints, the (discrete) 1-form dN acts on line segments connecting these midpoints[1]. For a triangle f with edges E_0, E_1, E_2 we denote the corresponding edge normals by N_0, N_1, N_2 and the connecting line segments by E_{ij}, *i.e.* E_{ij} connects the midpoint of E_i with that of E_j, *cf.* Fig. 1. In particular, we have the vector identity $E_k = -2\,E_{ij}$, where k is the complementary index to i and j in f. Using this notation, the fundamental theorem of calculus implies

$$dN(E_k) = -2\,dN(E_{ij}) = -2\int_{E_{ij}} dN = -2\,(N_j - N_i) = 2\,(N_i - N_j)\,. \tag{1.21}$$

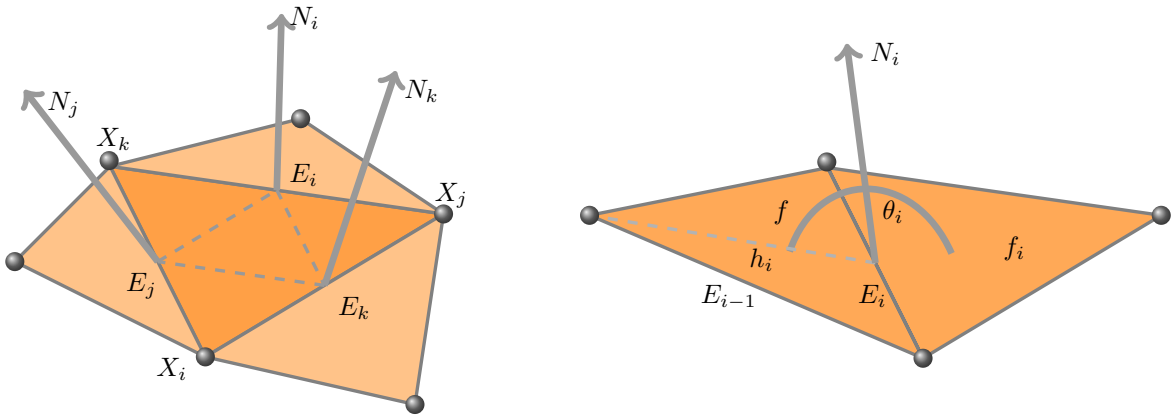


Figure 1: Support of discrete shape operator (left) and geometric interpretation of coefficients (right).

We can use (1.21) to simplify the entries of $H = H_f$ further, *i.e.*

$$\begin{aligned}
H_{11} &= \langle dN(E_2), E_2\rangle = 2\,\langle N_0 - N_1, E_2\rangle = 2\,\langle N_0, E_2\rangle + 2\,\langle N_1, E_0\rangle \\
H_{12} &= -\langle dN(E_2), E_1\rangle = -2\,\langle N_0 - N_1, E_1\rangle = 2\,\langle N_0, -E_1\rangle = 2\,\langle N_0, E_2\rangle \\
H_{21} &= -\langle dN(E_1), E_2\rangle = -2\,\langle N_2 - N_0, E_2\rangle = 2\,\langle N_0, E_2\rangle \\
H_{22} &= \langle dN(E_1), E_1\rangle = 2\,\langle N_2 - N_0, E_1\rangle = 2\,\langle N_0, E_2\rangle + 2\,\langle N_2, E_1\rangle
\end{aligned}$$

where we have used $\langle N_i, E_i\rangle = 0$ and $E_0 + E_1 + E_2 = 0$. Hence we get the representation

$$H = H_f = 2\sum_{i=0}^{2} \langle N_i, E_{i-1}\rangle\, M_i\,,$$

with a basis (M_0, M_1, M_2) of symmetric 2×2 - matrices given by

$$M_0 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad M_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

In the following, we will interprete the terms $\langle N_i, E_{i-1}\rangle$ geometrically. We refer to the height in f with base E_i by h_i, the neighboring triangle is denoted by f_i, *i.e.* $f \cap f_i = E_i$, *cf.* Fig. 1. Then by definition, the dihedral angle θ_i at E_i between f and f_i is given by two times the angle between h_i and N_i, since N_i is the angle bisector of θ_i by definition. We further use $h_i = -E_{i-1} + \beta E_i$ for some $\beta \in \mathbb{R}$ and obtain

$$\cos\frac{\theta_i}{2} = \langle N_i, \frac{h_i}{|h_i|}\rangle = -\frac{1}{|h_i|}\langle N_i, E_{i-1}\rangle + \beta\,\underbrace{\langle N_i, E_i\rangle}_{=0}\,.$$

[1] A continuous 1-form ω on a surface $\mathcal{S}$ is a mapping with $\omega(p) \in T_p^*\mathcal{S}$ for all $p \in \mathcal{S}$, where $T_p^*\mathcal{S}$ is the dual space of $T_p\mathcal{S}$. Continuous 1-forms are naturally evaluated as integrals along piecewise differentiable curves $\gamma : [a, b] \to \mathcal{S}$. An important example is given by $\omega = df$, where $f : \mathcal{S} \to \mathbb{R}$ is a differentiable function, and $\int_\gamma \omega = f(\gamma(b)) - f(\gamma(a))$. The concept of *Discrete Exterior Calculus (DEC)* aims at deriving a consistent theory of discrete forms on discrete manifolds, *i.e.* polygonal surfaces with certain properties. Analogously to the continuous setup, discrete 1-forms (*e.g.* given as a differential of a discrete function on a discrete manifold) are evaluated as integrals along discrete curves, *i.e.* polygonal chains. For further information and a comprehensive introduction on DEC we refer to [Hir03, DHLM05, DKT08].

Since $a_f := |f| = \frac{1}{2}|h_i| \, \|E_i\|$ we have $\langle N_i, E_{i-1} \rangle = -2 \frac{a_f}{\|E_i\|} \cos \frac{\theta_i}{2}$, hence

$$H_f = -4 \, a_f \sum_{i=0}^{2} \frac{\cos \frac{\theta_i}{2}}{\|E_i\|} \, M_i \, . \tag{1.22}$$

Finally, using (1.12), (1.22) and (1.3), we get a matrix representation of our discrete shape operator

$$S_f = G_f^{-1} H_f \in \mathbb{R}^{2,2} \, . \tag{1.23}$$

Based on the discrete shape operator, we can now derive a representation of discrete mean curvature. Analogously to the continous setting, the discrete mean curvature is defined as $\operatorname{tr} S_f$, hence it is also constant on faces. First, using (1.12) and (1.20) we get

$$G_f^{-1} = \frac{1}{\det G_f} \begin{pmatrix} \|E_1\|^2 & \langle E_1, E_2 \rangle \\ \langle E_1, E_2 \rangle & \|E_2\|^2 \end{pmatrix}$$

and due to (1.13) we have $\det G_f = 4a_f^2$ and hence one can easily show $\operatorname{tr}(G_f^{-1} M_i) = \frac{\|E_i\|^2}{4 \, a_f^2}$ for $i = 0, 1, 2$. Finally, this yields

$$\operatorname{tr} S_f = \operatorname{tr}(G_f^{-1} H_f) = -\sum_{i=0}^{2} \frac{\cos \frac{\theta_i}{2}}{a_f} \|E_i\| \, . \tag{1.24}$$

Remark (Evaluation of embedded discrete shape operator): If we consider the tangent space given by the plane of $T(f)$ the canonical basis is given by (1.20). Then, by definition, the matrix representation of the embedded discrete shape operator wrt. that basis is given by $S_f \in \mathbb{R}^{2,2}$.

References: The triangle-averaged discrete shape operator has been derived in joint work with Peter Schröder (Caltech), Max Wardetzky (Göttingen) and Benedikt Wirth (Münster). For further reading we refer to Behrend's diploma thesis [Hee11] resp. PhD thesis [Hee16] as well as to the corresponding publications [HRWW12, HRS$^+$14].

Generalization Grinspun *et al*. [GGRZ06] define a discrete shape operators on general meshes similar to the triangle-averaged operator introduced above. However, instead of prescribing an edge normal N_E to be the angle-bisecting normal at edge $E = T(f) \cap T'(f')$ the edge normal is supposed to fulfill $N_E \perp E$ and is parametrized over the angle γ_E between N_E and N_f. The vector $(\gamma_E)_E$ is then considered as another set of degrees of freedom.

2 Deformations of discrete shells

Eventually, we want to study deformations and deformation paths of discrete surfaces, *i.e.* triangle meshes. In applications, *e.g.* in animation movies, the mesh represents (the boundary/skin of) a complex character and the deformation path is supposed to describe a natural and non-trivial motion of that character. To obtain intuitive, visually appealing and natural results one needs to have a physically sound model. To this end, we go one step back and start the physical modeling in the continuous setup. Discrete surfaces are approximations of regular surfaces which have the physical interpretation of a *thin shell*. Vice versa, thin shells are three-dimensional solids with a high ratio from width to thickness. Mathematically, they are represented as compact embedded surface describing the midsurface of the material. Finally, this (regular/smooth) midsurface is approximated by a discrete surface.

Usually, one starts with theory of $3D$ elasticity and investigates deformation energies of some solid $\Omega_\delta \subset \mathbb{R}^3$ with δ being a tiny but finite thickness of the material, *i.e.* Ω_δ is a thin shell or plate from the physical point of view. Then on considers $\delta \to 0$ based on a suitable notion of convergence or by further apriori assumptions.

2.1 Elasticity theory

First, we give a survey on the theory of elastic deformations of solid three-dimensional bodies; for further reading we refer to [Cia88, MH94, Bra07]. Next, we derive a physically sound model for thin plates and shells based on concepts from $3D$ elasticity. This summary is based on the comprehensive and detailed descriptions found in several works by Ciarlet and co-workers [Cia00, Cia05, CM08].

Three-dimensional elasticity. Let $\mathcal{O} \subset \mathbb{R}^3$ be a homogenous[2] and solid object with boundary and $\phi \in W^{1,2}(\mathcal{O}; \mathbb{R}^3)$ a potentially large and nonlinear deformation. Typically, one assumes that ϕ is orientation preserving, *i.e.* $\det D\phi(x) > 0$ for all $x \in \mathcal{O}$, and injective (*i.e.*, no interpenetration of matter occurs). We postulate the existence of an *elastic deformation energy* $\mathcal{W}[\phi, \mathcal{O}]$ associated with the deformation ϕ. By definition, *elastic* means that $\mathcal{W}$ solely depends on the Jacobian $D\phi$ of ϕ. Furthermore, for so-called *hyperelastic* materials, $\mathcal{W}[\phi, \mathcal{O}]$ is the integral of an elastic energy density $W = W(D\phi)$, *i.e.*

$$\mathcal{W}[\phi, \mathcal{O}] = \int_{\mathcal{O}} W(D\phi) \, \mathrm{d}x \,. \tag{2.1}$$

A fundamental axiom of continuum mechanics is *frame indifference*, *i.e.* the invariance of the deformation energy with respect to rigid body motions. Hence, any coordinate transform $x \mapsto Qx + b$ for a rotation $Q \in SO(3)$ and a shift $b \in \mathbb{R}^3$ does not change the energy, *i.e.*

$$W(D\phi) = W(Q^{\mathrm{T}} D\phi Q) \quad \forall Q \in SO(3) \,.$$

A direct consequence is that W only depends on the so-called right Cauchy–Green strain tensor $C[\phi] = D\phi^{\mathrm{T}} D\phi$, which geometrically represents the metric measuring the deformed length in the undeformed reference configuration. The *elastic strain* is then defined by the difference $E[\phi] = \frac{1}{2}(C[\phi] - \mathbb{1})$.

Furthermore, we might assume $\mathcal{O}$ to be an *isotropic* material, *i.e.* a rotation of the material before applying a deformation yields the same energy as before, *i.e.*

$$W(D\phi) = W(D\phi Q) \quad \forall Q \in SO(3) \,.$$

It follows from the Rivlin-Erikson-Theorem [RE55] that the above two conditions lead to the fact that the energy density W only depends on the singular values $\lambda_1, \lambda_2, \lambda_3$ of $D\phi$, the so-called principal stretches. Instead of the principal stretches, one can equivalently describe the local deformation using the so-called invariants of the deformation gradient,

$$\begin{aligned} I_1 &= \|D\phi\|_F = \sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2} \,, \\ I_2 &= \|\mathrm{cof}\, D\phi\|_F = \sqrt{\lambda_1^2\lambda_2^2 + \lambda_1^2\lambda_3^2 + \lambda_2^2\lambda_3^2} \,, \\ I_3 &= \det D\phi = \lambda_1\lambda_2\lambda_3 \,, \end{aligned}$$

[2] This will later result in energy densities that do not depend on $x \in \mathcal{O}$.

where $\|A\|_F = \sqrt{\operatorname{tr}(A^{\mathrm{T}}A)}$ for $A \in \mathbb{R}^{d,d}$ and the cofactor matrix is given by $\operatorname{cof} A = \det A A^{-\mathrm{T}}$ for $A \in GL(d)$. Hence there is a function $\hat{W} : \mathbb{R}^3 \to \mathbb{R}$ with $W(D\phi) = \hat{W}(I_1, I_2, I_3)$, where I_1, I_2, and I_3 can be interpreted as the locally averaged change of an infinitesimal length, area, and volume during the deformation, respectively.

We shall furthermore assume that isometries, *i.e.* deformations with $D\phi^{\mathrm{T}} D\phi = \mathbb{1}$, are local minimizers with $W(D\phi) = 0$. Typical energy densities in this class are given by

$$W(D\phi) = \hat{W}(I_1, I_2, I_3) = a_1 I_1^p + a_2 I_2^q + \Gamma(I_3)\,, \tag{2.2}$$

for $a_1, a_2 > 0$ and a convex function $\Gamma : [0, \infty) \to \mathbb{R}$ with $\Gamma(I_3) \to \infty$ for $I_3 \to 0$ or $I_3 \to \infty$. In this work we focus on $p = q = 2$ which corresponds to the Mooney–Rivlin model [Cia88]. The built-in penalization of volume shrinkage, *i.e.* $\hat{W}(I_1, I_2, I_3) \to \infty$ for $\det D\phi \to 0$, enables us to control local injectivity. Incorporation of such a nonlinear elastic energy allows to describe large deformations with strong material and geometric nonlinearities, which cannot be treated by a linear elastic approach. A particular choice for a nonlinear elastic energy density was introduced in [Wir09] (*cf*. appendix A.1 of [WBRS11]) as

$$W(D\phi) = \frac{\mu}{2}\|D\phi\|_F^2 + \frac{\lambda}{4}(\det D\phi)^2 - \left(\mu + \frac{\lambda}{2}\right)\log\det D\phi - \frac{d\mu}{2} - \frac{\lambda}{4}\,, \qquad d \in \{2, 3\}\,. \tag{2.3}$$

Remark: The nonlinear elastic energy density (2.3) satisfies the following consistency condition:

$$W_{,FF}(\mathbb{1})(G, G) = \lambda(\operatorname{tr} G)^2 + \frac{\mu}{2}\operatorname{tr}\left((G + G^T)^2\right).$$

Furthermore, (2.3) is invariant with respect to rigid body motions, *i.e.* $W(D\phi) = 0$ iff. $\phi(x) = Qx + b$ with $Q \in SO(3)$.

Variational setup. In elasticity theory one typically considers variational problems as the minimization of

$$\mathcal{E}[\phi] = \int_{\mathcal{O}} W(D\phi)\,\mathrm{d}x - \int_{\mathcal{O}} F(\phi(x))\,\mathrm{d}x - \int_{\Gamma} G(\phi(x))\,\mathrm{d}a\,, \tag{2.4}$$

subject to suitable boundary conditions, where F and G represent body and boundary forces, respectively, and $\Gamma \subset \partial\mathcal{O}$. A corresponding existence theory for hyperelastic materials whose corresponding energy density W fulfills certain properties was established by John Ball [Bal77].

However, we will utilize $\mathcal{W}$ to define an elastic *dissimilarity measure* between *shapes*. That means, given two shapes $\mathcal{S}_A$ and $\mathcal{S}_B$ which are supposed to describe two elastic materials $\mathcal{O}_A$ and $\mathcal{O}_B$, we aim at minimizing $\phi \mapsto \mathcal{W}[\phi, \mathcal{S}_A]$ subject to the constraint $\phi(\mathcal{S}_A) = \mathcal{S}_B$. The dissimilarity measure is then given by

$$\mathrm{d}^2_{\text{elast}}(\mathcal{S}_A, \mathcal{S}_B) = \min_{\phi:\phi(\mathcal{S}_A)=\mathcal{S}_B} \int_{\mathcal{S}_A} W(D\phi)\,\mathrm{d}x\,. \tag{2.5}$$

Towards a two-dimensional theory. Physically, a shell is a three-dimensional body which is very thin in one dimension. The main objective of elastic shell theory is to predict the stress and the displacement arising in a thin shell in response to given forces. In a variational setup, this prediction is made by minimizing a suitable energy functional. In the following, we show how a simplification of the variational setup in three-dimensional elasticity (introduced above) leads to a two-dimensional theory. This simplification is done by exploiting the special geometry of the shell, and especially, the assumed "smallness" of the thickness of the shell, denoted by $\delta > 0$, *cf*. Fig. 2. Eventually, this assumption allows to eliminate some of the terms of lesser order of magnitude with respect to the thickness of the shell.

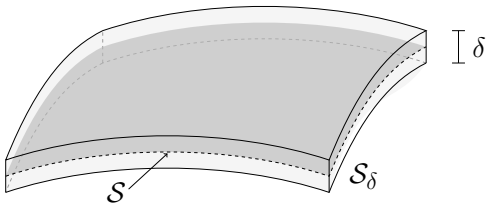


Figure 2: Elastic shell $\mathcal{S}_\delta \subset \mathbb{R}^3$ with finite thickness $\delta > 0$ and midsurface $\mathcal{S}$.

For simplicity, we will first consider *plate theories*, where the undeformed/reference configuration is *flat*. In contrast, in *shell theories*, the undeformed/reference configuration is already curved, *i.e.* given by some material

with midsurface $\mathcal{S} \subset \mathbb{R}^3$. Let $\omega \subset \mathbb{R}^2$ be a domain in the plane, $\delta > 0$, and $\Omega_\delta = \omega \times (-\frac{\delta}{2}, \frac{\delta}{2}) \subset \mathbb{R}^3$ be a thickened plate, $p_\delta = (\xi, z) \in \omega \times (-\frac{\delta}{2}, \frac{\delta}{2})$. A deformation $\phi_\delta \in H^1(\Omega_\delta, \mathbb{R}^3)$ of this plate is characterized by a stored energy function $\mathcal{W}$ as in (2.1), *i.e.*

$$\mathcal{W}[\Omega_\delta, \phi_\delta] = \int_{\Omega_\delta} W(D\phi_\delta) \, \mathrm{d}p_\delta \,,$$

and the functional $\mathcal{W}[\Omega_\delta, \phi_\delta]$ is optimized subject to forces and boundary conditions.

In the *two-dimensional approach to shell theory*, the above minimization problem is replaced by a (presumably simpler) two-dimensional problem, which is eventually posed over the middle layer $\omega \subset \mathbb{R}^2$ (resp. middle surface $\mathcal{S} \subset \mathbb{R}^3$) of the plate (resp. shell). The two-dimensional approach to shell theory yields a variety of different shell models, which can be classified into two categories:

(i) The first category of models is obtained from the three-dimensional problem formulation by letting the thickness $\delta > 0$ of the shell go to zero. This can be formulated rigorously by means of Γ-convergence [DGDM83, Bra02]. Depending on the scaling, boundary conditions and applied forces one obtains either a so-called *membrane* shell model [LDR95, LDR96], or a flexual or *bending* shell model [FJM02a, FJM02b, FJMM03].

(ii) The second category of models are obtained from the three-dimensional model by restricting the range of admissible deformations by means of specific a priori assumptions that are supposed to take into account the smallness of the thickness. For example, the (geometric) Kirchhoff-Love assumptions [Kir50, Lov88], *i.e.* (a) any point situated on a normal to the middle surface remains on the normal to the deformed middle surface after the deformation has taken place and (b) that the distance between such a point and the middle surface remains constant, combined with mechanical assumption by John [Joh65], *i.e.* the state of stress inside the shell is planar and parallel to the middle surface, lead to Koiter's shell model [Koi66] (see also [RS10]). More general models of this category are of the Middlin-Reissner type, *e.g.* the Cosserat model, which does not postulate the Kirchhoff-Love assumption.

In the following, we will discuss the ansatz-free former concept (i) in detail.

2.2 Membrane and bending energies by Γ-convergence

In this section we discuss the ansatz-free approach by means of Γ-convergence. For an introduction to the concepts of Γ-convergence we refer to [DGDM83, Bra02]. As mentioned before we first study *plate theories* derived from $3D$ elasticity.

Remark: Let us emphasize in particular that we focus on the *qualitative* understanding of the rigorous derivations, *i.e.* we are interested in which objects the limit depends on. These objects will then be used to define a physically sound dissimilarity measure between two thin shells. Hence we neglect *quantitative* aspects, *e.g.* the detailed shape of limit integrands or values of physical constants in the limit, as these will later be chosen individually for different applications.

Let $\omega \subset \mathbb{R}^2$ bounded and simply connected, $\gamma := \partial\omega$ Lipschitz. As, we first only consider *plates* here, that means the two-dimensional domain ω is flat and not curved. For $\delta > 0$ we define the reference domain of a *thin plate* by $\Omega_\delta = \omega \times (-\frac{\delta}{2}, \frac{\delta}{2})$. The starting point is the elastic energy of a deformation $\phi_\delta \in H^1(\Omega_\delta, \mathbb{R}^3)$, *i.e.*

$$\mathcal{W}[\phi_\delta, \Omega_\delta] = \int_{\Omega_\delta} W(D\phi) \, \mathrm{d}p_\delta \tag{2.6}$$

for some frame-indifferent elastic energy density $W : \mathbb{R}^{3,3} \to \mathbb{R}$ which is minimized on $SO(3)$ and fulfills $W(\mathbb{1}) = 0$ and $W(F) = \infty$ if $\det F \leq 0$. Furthermore, one usually assumes some regularity and certain growth conditions. Typical energy densities are for instance

- The distance to the special orthogonal group, *i.e.* $W(F) = \mathrm{dist}^2(F, SO(3)) \approx \frac{1}{4}\|F^T F - \mathbb{1}\|_F^2$.

- A generic isotropic material is described by the *St Venant-Kirchhoff* density given by

$$W^{\text{StVK}}(F) = \frac{\lambda}{8}(\text{tr}\,(F^TF - \mathbb{1}))^2 + \frac{\mu}{4}\text{tr}\,(F^TF - \mathbb{1})^2\,. \tag{2.7}$$

  Note that $W^{\text{StVK}}(D\phi_\delta) = \frac{\lambda}{2}(\text{tr}\,E[\phi_\delta])^2 + \mu\text{tr}\,(E[\phi_\delta]^2)$, with $E[\phi_\delta] = \frac{1}{2}(D\phi_\delta^T D\phi_\delta - \mathbb{1})$.

- A general Mooney-Rivlin model:

$$W(F) = a\|F\|_F^2 + b\|\text{cof}\,F\|_F^2 + c(\det F)^2 - d\log\det F + e\,,$$

  with $e \in \mathbb{R}$ s.t. $W(\mathbb{1}) = 0$. Neo-Hooke material: $b = 0$.

In the following we will consider the limit behaviour of (2.6) for $\delta \to 0$.

An excursion in Γ-convergence.

Definition 2.1 (Γ-convergence)**.** Let (X, d) be a metric space, $F_j : X \to \mathbb{R}$ a sequence of functionals and $F : X \to \mathbb{R}$. Then F_j Γ-converges to F with respect to d, *i.e.* $F_j \xrightarrow{\Gamma} F$, if

(i) *liminf condition*. For every sequence $(x_j)_j \subset X$ with $d(x_j, x) \to 0$ for some $x \in X$ we have

$$F[x] \le \liminf_{j\to\infty} F_j[x_j]\,.$$

(ii) *limsup condition/recovery sequence.* For each $x \in X$ there is a sequence $(x_j)_j \subset X$ with $d(x_j, x) \to 0$ and

$$F[x] \ge \limsup_{j\to\infty} F_j[x_j]\,.$$

Some remarks on Γ-convergence:

(i) If $F_j \xrightarrow{\Gamma} F$ then F is lower semi-continuous (w.r.t. d), *i.e.* if $d(x_j, x) \to 0$ then $F[x] \le \liminf_{j\to\infty} F[x_j]$.

(ii) If $F_j \xrightarrow{\Gamma} F$ and G is continuous, then $F_j + G \xrightarrow{\Gamma} F + G$.

(iii) But, if $F_j \xrightarrow{\Gamma} F$ and $G_j \xrightarrow{\Gamma} G$ then *not* necessarily $F_j + G_j \xrightarrow{\Gamma} F + G$.

(iv) If $F_j = F$ for all j, then *not* necessarily $F_j \xrightarrow{\Gamma} F$. (Limit has to be lsc.!)

Note the importance of lsc. functions: The *direct method of Calculus of Variations* states that if X is a reflexive Banach space, F is weakly lower semi-continuous and coercive, then F attains its minimum on X.

Example: $X = \mathbb{R}$ with $d(x, y) = \|x - y\|$, $F_j(x) = \sin(jx)$ and $F(x) \equiv -1$. Then $F_j \xrightarrow{\Gamma} F$. The *liminf condition* is trivial. Let $x \in \mathbb{R}$. Define $x_j \in \mathbb{R}$ s.t. $d(x, x_j) = \min\{d(x, y) : \sin(jy) = -1\}$. Then $d(x, x_j) \to 0$ and $-1 \ge F_j(x_j)$, which proves the *limsup condition*.

A function F is said to be coercive if the following holds: If $F[x_j]$ is bounded, then $(x_j)_j$ is precompact in X, *i.e.* has a converging subsequence. A function F is mildly coercive if $\inf_{x\in X} F[x] = \inf_{x\in K} F[x]$ for some compact set $K \subset X$. A sequence $(F_j)_j$ is said to be *equi-coercive* if the following holds: If $F_j(x_j)$ is bounded then $(x_j)_j$ is precompact in X. A sequence $(F_j)_j$ is said to be *equi-mildly coercive* if there is some compact set $K \subset X$ such that $\inf_{x_j\in X} F_j[x_j] = \inf_{x_j\in K} F_j[x_j]$ for all j. If F is coercive, then it is mildly coercive.

Some properties related to the existence/computation of minimizers:

(i) Let $F_j \xrightarrow{\Gamma} F$, x_j minimizer of F_j and $x_j \to x$. Then x is minimizer of F and $F_j[x_j] \to F[x]$.

(ii) Let $F_j \xrightarrow{\Gamma} F$, $(F_j)_j$ equi-mildly coercive, then thre is a minimizer x of F with $F[x] = \lim_{j\to\infty} \inf_{x_j\in X} F_j[x_j]$. Moreover, if $(x_j)_j$ is minimizing sequence, then every accumulation point is a minimum point of F.

Proof (i): Let us assume there is $x' \in X$ with $F[x'] < F[x]$. Then there is a sequence $x'_j \to x'$ with

$$F[x'] \underset{\text{Def.2.1}(ii)}{\geq} \limsup_{j\to\infty} F_j[x'_j] \geq \liminf_{j\to\infty} F_j[x'_j] \underset{x_j \text{ min. of } F_j}{\geq} \liminf_{j\to\infty} F_j[x_j] \underset{\text{Def.2.1}(i)}{\geq} F[x] \quad \lightning$$

Remark I: A minimizing sequence $(x_j)_j$ fulfills $\lim_j F_j[x_j] = \lim_j \inf_{x\in X} F_j[x]$ which is *not* necessarily a sequence of minimizers (which might not exist!).
Remark II: Note that (i) requires existence of minimizers of F_j *and* convergence of minimizers, whereas (ii) provides existance of minimizers without these two conditions.

Applications of Γ-convergence are for instance phase transitions or dimension reductions. In the latter case we consider the functional (2.6) depending on a small parameter $\delta > 0$ with $\delta \to 0$. However, for each $\delta > 0$ the functional is defined on a different function space, *i.e.* $H^1(\Omega_\delta, \mathbb{R}^3)$. Hence one considers the variable transformation $p_\delta = (\xi, z) \in \Omega_\delta \mapsto p = (\xi, \delta^{-1}z) \in \Omega_1$ and the corresponding gradient transformation $D_\delta = (D_\xi, \frac{1}{\delta}\partial_z)$, and defines a rescaled functional by

$$\mathcal{W}^*[\phi_\delta, \Omega_1] = \delta \int_{\Omega_1} W(D_\delta \phi_\delta(p))\,\mathrm{d}p \tag{2.8}$$

where $p = (\xi, z)$. Note that still $\mathcal{W}^*[\phi_\delta, \Omega_1] = \mathcal{W}[\phi_\delta, \Omega_\delta]$.

Depending on the boundary conditions a characteristic scaling of the elastic energy (2.6) resp. (2.8) is observed. If the boundary conditions induce a stretching of the midplane ω, *e.g.* by

$$\Omega_\delta = (-1,1)^2 \times (-\frac{\delta}{2}, \frac{\delta}{2}), \quad \phi_\delta(p_\delta)\Big|_{\xi_1=\pm 1} = p_\delta \pm (a,0,0)\,,$$

for some $a > 0$, we observe $\mathcal{W}^*[\phi_\delta, \Omega_1] \sim \mathcal{W}[\phi_\delta, \Omega_\delta] \sim \delta$. If we apply compressive boundary conditions, we have

$$\Omega_\delta = (-1,1)^2 \times (-\frac{\delta}{2}, \frac{\delta}{2}), \quad \phi_\delta(p_\delta)\Big|_{\xi_1=\pm 1} = p_\delta \mp (b,0,0)\,, \tag{2.9}$$

for some $b \in (0,1)$, we observe $\mathcal{W}^*[\phi_\delta, \Omega_1] \sim \mathcal{W}[\phi_\delta, \Omega_\delta] \sim \delta^3$. In particular, the plate will accommodate the boundary conditions by bending while keeping its midsurface unstretched. This leads to the investigation of two different types of models. First, a *membrane limit theory* is derived by studying the Γ-convergence of the rescaled functional $\delta^{-1}\,\mathcal{W}^*[\phi_\delta, \Omega_1]$ for $\delta \to 0$. Second, a *bending limit theory* is derived by studying the Γ-convergence of the rescaled functional $\delta^{-3}\,\mathcal{W}^*[\phi_\delta, \Omega_1]$ for $\delta \to 0$.

Γ-limit for the membrane model. We assume the following growth condition

$$c_1\|F\|^2 - c_2 \leq W(F) \leq c_3\|F\|^2 + c_4\,, \tag{2.10}$$

for real numbers $c_i \geq 0$ and define rescaled functional

$$\mathcal{W}^\delta_{\text{mem}}[\phi_\delta, \Omega_\delta] := \frac{1}{\delta}\mathcal{W}^*[\phi_\delta, \Omega_1] = \int_{\Omega_1} W(D_\delta \phi(p))\,\mathrm{d}p\,.$$

Theorem 2.2 (Membrane Γ-limit, [LDR95, LDR96])**.** *The sequence $(\mathcal{W}^\delta_{mem})_\delta$ is uniformly coercive in H^1 and $\mathcal{W}^\delta_{mem} \xrightarrow{\Gamma} \mathcal{W}_{mem}$ w.r.t. the weak H^1-topology with*

$$\mathcal{W}_{mem}[\phi, \omega] = \int_\omega QW_{2D}(D\phi)\,\mathrm{d}\xi\,, \quad \phi \in H^1(\omega, \mathbb{R}^3)\,,$$

where $QW_{2D} : \mathbb{R}^{3,2} \to \mathbb{R}$ arises from a double relaxation process:

(1) For $F \in \mathbb{R}^{3,2}$ define $W_{2D}(F) = \min_{b\in\mathbb{R}^3} W(F|b)$

(2) Computation of quasi-convex envelope, i.e. *$QW_{2D}(F) = \inf\{\fint_\omega W_{2D}(F + D\phi(x))\,\mathrm{d}x : \phi \in W^{1,\infty}_0\}$*

If W fulfills the growth condition (2.10), then the relaxation of $\int_\Omega W(D\phi)\,\mathrm{d}x$ is given by $\int_\Omega QW(D\phi)\,\mathrm{d}x$ (w.r.t. the weak H^1-topology). The relaxation $\mathrm{rel}\,F$ of some functional F is given by $\mathrm{rel}\,F = \sup_G\{G \text{ lsc. } : G \le F\}$.

Some remarks:

- Thm.2.2 holds in $W^{1,p}$ for $p \in (1,\infty)$ if (2.10) is replaced by a corresponding p-growth condition.
- A corresponding result holds for a curved reference domain, *i.e.* ω is replaced by some compact and smooth surface $\mathcal{S} \subset \mathbb{R}^3$.
- For general elastic densities W the computation of QW_{2D} can be very complex.

The growth conditions (2.10) are fulfilled by typical energy densities representing isotropic materials, *e.g.* by (2.7). Furthermore, if the original density W was frame-indifferent, then the limit density is also frame-indifferent and depends on the metric of the deformed middle surface only [LDR96]. If additionally $W(F) \ge W(\mathbb{1})$ for all $F \in \mathbb{R}^{3,3}$, which is always the case in our examples as we assume $W(\mathbb{1}) = 0$ and $W \ge 0$, the corresponding membrane shell energy is constant under compression, *i.e.* the shell offers no resistance to crumpling [LDR96].

Qualitative properties of membrane limit. For frame-indifferent and isotropic densities the limit membrane energy can be written as an integral over the midsurface $\mathcal{S}$ whose integrand depends on the principal invariants of the right Cauchy-Green strain tensor $C[\phi] = D\phi^T D\phi$ only (*cf.* Sec. 2.1), where $\phi : \mathcal{S} \to \mathbb{R}^3$ is a deformation of the midsurface $\mathcal{S}$ to $\mathcal{S}^\phi = \phi(\mathcal{S})$. In detail, the pointwise linear operator $C[\phi]$ measures the distortion of tangent vectors which are mapped from $T_p\mathcal{S}$ to $T_{\phi(p)}\mathcal{S}^\phi$ for some arbitrary point $p \in \mathcal{S}$, *i.e.*

$$g_p(C[\phi]V, W) = g_{\phi(p)}(D\phi\,V, D\phi\,W)\,, \quad V, W \in T_p\mathcal{S}\,. \tag{2.11}$$

If we assume that a local neighborhood of p and $\phi(p)$, respectively, are parametrized over the same domain $\omega \subset \mathbb{R}^2$ by immersions $x, x^\phi : \omega \to \mathbb{R}^3$, we can formally write $\phi = x^\phi \circ x^{-1}$ and hence $D\phi = Dx^\phi(Dx)^{-1}$. This concatenation property has been used in [CLR04, LDRS05] to derive a two-dimensional representation of $C[\phi] \in \mathbb{R}^{3,3}$ by a distortion tensor $\mathcal{G}[\phi] \in \mathbb{R}^{2,2}$. In detail, since $Dx^T \cdot Dx g^{-1} = \mathbb{1}_{2,2}$ we have

$$C[\phi] = D\phi^T D\phi = (Dx)^{-T} g^\phi (Dx)^{-1} = Dx g^{-1} g^\phi (Dx)^{-1}\,, \tag{2.12}$$

with $g = Dx^T Dx$ and $g_\phi = (Dx^\phi)^T(Dx^\phi)$ denoting the first fundamental form of the undeformed and deformed configuration, respectively. Let $v \in T_\xi\omega$. If we apply $Dxv \in T_p\mathcal{S}$ to both sides of (2.12) we get

$$C[\phi]\,Dxv = Dx\left(g^{-1}g^\phi v\right)\,,$$

which leads to the pointwise definition

$$\mathcal{G}[\phi] = g^{-1} g_\phi \tag{2.13}$$

From the considerations above we deduce a membrane shell energy $\mathcal{W}_{\text{mem}}$ which is supposed to measure the dissimilarity in terms of tangential stretching and shearing induced by a deformation ϕ of the undeformed (reference) shell $\mathcal{S}$, *i.e.*

$$\mathcal{W}_{\text{mem}}[\mathcal{S}, \phi] = \int_{\mathcal{S}} W_{\text{mem}}(\mathcal{G}[\phi])\,\mathrm{d}a\,. \tag{2.14}$$

We shall make use of the density defined in (2.3) with $d = 2$. In particular, we have $W_{\text{mem}}(F) = W_{\text{mem}}(\mathrm{tr}\,F, \det F)$ as well as $W_{\text{mem}}(\mathbb{1}) = 0$ and $\partial_F W_{\text{mem}}(\mathbb{1}) = 0$.

Γ-limit for the bending model (for plates). We have seen that the elastic energy (2.6) resp. (2.8) scales like δ^3 if we apply compressive boundary conditions (2.9). In particular, the plate accommodates the boundary conditions by bending while keeping its midsurface unstretched. However, as the volume of $\mathcal{S}_\delta$ scales like δ the integrand $W(D\phi_\delta)$ approaches zero much faster. That means, since W is assumed to be minimized exactly on $SO(3)$, the Jacobian $D\phi_\delta \in \mathbb{R}^{3,3}$ tends in a certain sense to $SO(3)$. Friesecke, James and Müller [FJM02b] came up with a rigorous derivation of the thin-plate limit of three-dimensional nonlinear elasticity theory, not just under the special compressive boundary conditions considered above but under any boundary condition that does not induce tangential distortion of the midsurface. As for the derivation of the membrane model, the mathematical setting in

which these results are formulated is that of Γ-convergence. However, due to the scaling mentioned above, for the derivation of the bending model the limit process of

$$\mathcal{W}^{\delta}_{\text{bend}}[\phi_\delta, \Omega_\delta] := \frac{1}{\delta^3} \int_{\Omega_\delta} W(D\phi_\delta) \, \mathrm{d}p_\delta = \frac{1}{\delta^2} \int_{\Omega_1} W(D_\delta \phi_\delta) \, \mathrm{d}p \tag{2.15}$$

is considered for $\delta \to 0$. Although the approaches are similar, the bending model is more difficult to derive since the limit functional contains higher derivatives and one is thus dealing with a singular perturbation problem.

Let $W : R^{3,3} \to \mathbb{R}$ be a continuous and frame-indifferent energy density fulfilling the growth condition $W(F) \geq c \operatorname{dist}^2(F, SO(3))$ as well as $W(F) = 0$ if $F \in SO(3)$. For simplicity, we further assume that W represents an *isotropic* material which satisfies the consistency relation

$$W_{,FF}(\mathbb{1})(G, G) = \lambda(\operatorname{tr} G)^2 + \frac{\mu}{2} \operatorname{tr}\left((G + G^T)^2\right). \tag{2.16}$$

For example, W describes a St Venant-Kirchhoff material as in (2.7).

Theorem 2.3 (Bending Γ-limit for plates, [FJM02a, FJM02b])**.** *Under the assumptions on W stated above, the following convergence holds in the H^1-topology for $\delta \to 0$:*

$$\mathcal{W}^{\delta}_{\textit{bend}}[\phi_\delta, \Omega_\delta] \quad \xrightarrow{\Gamma} \quad \mathcal{W}^{0}_{\textit{plate}}[\phi, \omega] \,,$$

where $\mathcal{W}^{\delta}_{\textit{bend}}[\phi_\delta, \Omega_\delta]$ as in (2.15) *and*

$$\mathcal{W}^{0}_{\textit{plate}}[\phi, \omega] = \begin{cases} \frac{1}{24} \int_\omega \left(2\mu \operatorname{tr}\left(h[\phi]^2\right) + \frac{\lambda\mu}{\mu + \lambda/2} (\operatorname{tr} h[\phi])^2\right) \mathrm{d}\xi & , \textit{on isometries } \phi : \omega \to \mathbb{R}^3 \\ +\infty & , \textit{otherwise} \end{cases}.$$

Remark: The limiting energy thus depends on the second fundamental form $h[\phi] = Dn^T D\phi$, where ϕ can be thought of being a parametrization of the deformed plate $\phi(\omega)$, *i.e.* $n \parallel (\phi_{,1} \times \phi_{,2})$. Note that $\phi : \omega \to \mathbb{R}^3$ is an isometry iff. $(g_\phi)_{ij} = \phi_{,i} \cdot \phi_{,j} = \delta_{ij}$, *i.e.* in particular $\det g_\phi = 1$.

**Idea of recovery sequence (*cf. *[Bar15]).* For simplicity, we consider a St.Venant-Kirchhoff material (2.7) with $\lambda = 0$ and $\mu = 1$, *i.e.* $W(F) = \frac{1}{4}\|F^T F - \mathbb{1}\|^2$. Furthermore, we consider the representation (2.8), *i.e.* $\mathcal{W}^{\delta}_{\text{bend}}[\phi_\delta, \Omega_\delta] = \delta \int_{\Omega_1} W(D_\delta \phi_\delta) \, \mathrm{d}p_\delta$, with $D_\delta = (D_\xi, \frac{1}{\delta}\partial_z)$. Let $\phi : \omega \to \mathbb{R}^3$ a deformation of the middle plate that we seek to *recover*. We assume the sequence of deformations $\phi_\delta : \Omega_1 \to \mathbb{R}^3$ is of the form

$$\phi_\delta(\xi, z) = \phi(\xi) + \delta z n(\xi) \,,$$

where n is the unit normal to the surface parametrized by ϕ, *i.e.* $\langle \partial_k \phi(\xi), \, n(\xi) \rangle = 0$ for $k = 1, 2$. This means, segments normal to ω are mapped to straight lines that are normal to the deformed surface (*cf*. Kirchhoff-Love hypothesis). We consider the splitting

$$D_\delta \phi_\delta = [D_\xi \phi, n] + \delta z [D_\xi n, 0] \,.$$

Since the second fundamental form $h = h[\phi] = (D_\xi \phi)^T D_\xi n$ is symmetric and $(D_\xi n)^T n = 0$ we get

$$\begin{aligned} D_\delta \phi_\delta^T D_\delta \phi_\delta &= \begin{bmatrix} (D_\xi \phi)^T D_\xi \phi & 0 \\ 0 & |n|^2 \end{bmatrix} + \delta z \begin{bmatrix} 2 D_\xi \phi^T D_\xi n & (D_\xi n)^T n \\ n^T D_\xi n & 0 \end{bmatrix} + \delta^2 z^2 \begin{bmatrix} (D_\xi n)^T D_\xi n & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} g + 2\delta z h + \delta^2 z^2 r & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

where $g = g[\phi] = (D_\xi \phi)^T D_\xi \phi$ is the first fundamental form and $r := (D_\xi n)^T D_\xi n$. Hence

$$\begin{aligned} \mathcal{W}^{\delta}_{\text{bend}}[\phi_\delta, \Omega_\delta] &= \frac{1}{\delta^3} \cdot \delta \int_{\Omega_1} \frac{1}{4} \|D_\delta \phi_\delta^T D_\delta \phi_\delta - \mathbb{1}\|^2 \, \mathrm{d}p = \frac{1}{4\delta^2} \int_\omega \int_{-\frac{1}{2}}^{\frac{1}{2}} \|(g - \mathbb{1}) + 2\delta z h + \delta^2 z^2 r\|^2 \, \mathrm{d}z \, \mathrm{d}\xi \\ &= \frac{1}{4\delta^2} \int_\omega \int_{-\frac{1}{2}}^{\frac{1}{2}} \|g - \mathbb{1}\|^2 + 4\delta z (g - \mathbb{1}) : h + 2\delta^2 z^2 (g - \mathbb{1}) : r + 4\delta^2 z^2 \|h\|^2 + 4\delta^3 z^3 h : r + \delta^4 z^4 \|r\|^2 \, \mathrm{d}z \, \mathrm{d}\xi \\ &= \int_\omega \frac{1}{4\delta^2} \|g - \mathbb{1}\|^2 + \frac{1}{24} (g - \mathbb{1}) : r + \frac{1}{12} \|h\|^2 + \frac{\delta^2}{20} \|r\|^2 \, \mathrm{d}\xi \\ &\xrightarrow{\delta \to 0} \begin{cases} \frac{1}{12} \int_\omega \|h\|^2 \, \mathrm{d}\xi & , \|g - \mathbb{1}\|^2 = 0 \\ +\infty & , \text{otherwise} \end{cases}, \end{aligned}$$

which correponds to $\mathcal{W}^0_{\text{plate}}[\phi,\omega]$ in Thm.2.3 with $\lambda=0$ and $\mu=1$.

Γ-limit for the bending model (for shells). More general, the two-dimensional midsurface of the reference configuration is already curved. Instead of $\omega\subset\mathbb{R}^2$ and $\Omega_\delta=\omega\times(-\delta/2,\delta/2)$, respectively, we consider a smooth and compact surface $\mathcal{S}\subset\mathbb{R}^3$ and

$$\mathcal{S}_\delta=\left\{x+z\,n(x)\,\middle|\,x\in\mathcal{S},z\in\left(-\frac{\delta}{2},\frac{\delta}{2}\right)\right\}.$$

Theorem 2.4 (Bending Γ-limit for shells, [FJMM03])**.** *Under similar assumptions as in Thm.2.3 the scaled energy*

$$\mathcal{W}^\delta_{\textit{bend}}[\phi_\delta,\mathcal{S}_\delta]=\frac{1}{\delta^3}\int_{\mathcal{S}_\delta}W(D\phi_\delta)\,\mathrm{d}p_\delta$$

Γ-converges w.r.t the H^1-topology to a two-dimensional limit functional given by

$$\mathcal{W}^0_{\textit{shell}}[\phi,\mathcal{S}]=\begin{cases}\frac{1}{24}\int_{\mathcal{S}}\min_{v\in\mathbb{R}^3}Q(S^{\textit{rel}}_\phi(p)+v\otimes n(p))\,\mathrm{d}a & ,\phi\in\mathcal{A}\\ +\infty & ,\textit{otherwise}\end{cases},$$

with the quadratic form $Q(G)=W_{,FF}(\mathbb{1})(G,G)$ and the admissible set of isometric deformations

$$\mathcal{A}=\{\phi\in W^{2,2}(\mathcal{S},\mathbb{R}^3)\,|\,(D_{\textit{tan}}\phi)^T(D_{\textit{tan}}\phi)=\mathbb{1}\ \textit{a.e. on}\ \mathcal{S}\}.$$

Here the tangential derivative $D_{\tan}\phi\in\mathbb{R}^{3,2}$ can be extended to a proper rotation $Q(p)=Q[\phi](p)\in SO(3)$ if ϕ is isometric. The two-dimensional limit energy density depends on the *relative shape operator* $S^{\text{rel}}_\phi(p):T_p\mathcal{S}\to T_p\mathcal{S}$ as it has been defined in Def. 1.9. Note that the limit bending energy $\mathcal{W}^0_{\text{shell}}[\phi,\mathcal{S}]$ is only finite for deformations $\phi\in\mathcal{A}$, hence we will assume in the remainder of this paragraph that we are dealing with isometric deformations.

Remark: In Sec. 1.3 the relative shape operator has been defined via the pulled-back shape operator, *cf*. Def. 1.8. A different (but equivalent) derivation is given by the pointwise definition

$$S^{\text{rel}}_\phi(p)=S(p)-Q(p)^T\,S_\phi(\phi(p))\,Q(p)\,,\quad p\in\mathcal{S},\tag{2.17}$$

where $S(p):T_p\mathcal{S}\to T_p\mathcal{S}$ and $S_\phi(q):T_q\mathcal{S}^\phi\to T_q\mathcal{S}^\phi$ are the shape operators on the undeformed and deformed configuration, respectively. Here we have used the notation $\mathcal{S}^\phi=\phi(\mathcal{S})$ and $q=\phi(p)$ for $p\in\mathcal{S}$. The relative shape operator is supposed to measure the (pointwise) difference between the shape operators on $\mathcal{S}$ and $\mathcal{S}^\phi$, respectively. However, as these operators live on different tangent spaces, *i.e*. rotated planes in $\mathbb{R}^3$, we must include proper rotations to ensure well-definedness of the pointwise difference. Hence $Q(p)$ and $Q(p)^T$ denote the linear mappings between the two different tangent spaces, as illustrated in the following diagram:

$$\begin{array}{ccc}T_p\mathcal{S} & \xrightarrow{\quad S(p)\quad} & T_p\mathcal{S}\\ \Big\downarrow{\scriptstyle Q(p)} & & \Big\uparrow{\scriptstyle Q(p)^T}\\ T_{\phi(p)}\phi(\mathcal{S}) & \xrightarrow[\quad S_\phi(\phi(p))\quad]{} & T_{\phi(p)}\phi(\mathcal{S})\end{array}$$

We have $Q(p)=D\phi(p)\in SO(3)$ and $D\phi(p)n(p)=n^\phi(\phi(p))$, where n^ϕ denotes the normal on the deformed surface. To this end, we can think of $Q^T(S_\phi\circ\phi)Q$ as being a *pulled-back* representation S^*_ϕ of the shape operator S_ϕ on the deformed configuration. The linear operator $S^*_\phi:T_p\mathcal{S}\to T_p\mathcal{S}$ is then implicitly defined as in Def. 1.8.

As for the membrane shell energy we use the analytic results presented above to extract a generic bending shell energy by setting

$$\mathcal{W}_{\text{bend}}[\mathcal{S},\phi]=\int_{\mathcal{S}}W_{\text{bend}}(S^{\text{rel}}_\phi)\,\mathrm{d}a\,.\tag{2.18}$$

In general, we make use of the density

$$W_{\text{bend}}(A)=\alpha(\operatorname{tr}A)^2+(1-\alpha)\,\|A\|^2_F\,,\quad\alpha\in\{0,1\}\,.\tag{2.19}$$

In particular, for $\alpha = 1$ we recover an adapted form of the Willmore energy measuring differences in mean curvature. Recall that the matrix representation of the relative shape operator in the parameter domain was defined in (1.6)

$$s_\xi^{\text{rel}}[\phi] = s_\xi - s_\xi^*[\phi] = g_\xi^{-1}(h_\xi - \tilde{h}_\xi)\,.$$

It is not difficult to verify that for $\alpha = 0$ we get

$$\mathcal{W}_{\text{bend}}[\mathcal{S}, \phi] = \int_{\mathcal{S}} \|S_\phi^{\text{rel}}\|_F^2 \,\mathrm{d}a = \int_\omega \operatorname{tr}\left(s_\xi^{\text{rel}}[\phi]^2\right)\sqrt{\det g}\,\mathrm{d}\xi\,, \tag{2.20}$$

and for $\alpha = 1$ we get

$$\mathcal{W}_{\text{bend}}[\mathcal{S}, \phi] = \int_{\mathcal{S}} \left(\operatorname{tr} S_\phi^{\text{rel}}\right)^2 \mathrm{d}a = \int_\omega \left(\operatorname{tr} s_\xi^{\text{rel}}[\phi]\right)^2 \sqrt{\det g}\,\mathrm{d}\xi\,. \tag{2.21}$$

**Sketch of proof [Hee16]:* In the following we drop the specification of the point $p \in \mathcal{S}$ in the notation. We have

$$\|S_\phi^{\text{rel}}\|_F^2 = \|S - S_\phi^*\|_F^2 = \sum_{i,j=1}^2 \left[\langle e_i, Se_j\rangle_{\mathbb{R}^3} - \langle e_i, S_\phi^* e_j\rangle_{\mathbb{R}^3}\right]^2,$$

$$\operatorname{tr} S_\phi^{\text{rel}} = \operatorname{tr}(S - S_\phi^*) = \sum_{i=1}^2 \left[\langle e_i, Se_i\rangle_{\mathbb{R}^3} - \langle e_i, S_\phi^* e_i\rangle_{\mathbb{R}^3}\right],$$

where (e_1, e_2, e_3) is the canonical basis of $\mathbb{R}^3$. Let us assume that a neighboorhoud of $p \in \mathcal{S}$ is parametrized by some chart $x : \omega \subset \mathbb{R}^2 \to \mathbb{R}^3$. For $\xi \in \omega$ such that $p = x(\xi)$ we have another basis (v_1, v_2, n) with $[v_1|v_2] = Dx(\xi)$ and $n = n(p)$ with

$$e_i = a_{1i}v_1 + a_{2i}v_2 + a_{3i}n\,,\quad a_i := \begin{pmatrix} a_{1i} \\ a_{2i} \\ a_{3i}\end{pmatrix} = [v_1|v_2|n]^{-1}e_i =: Ae_i\,,$$

where $A \in \mathbb{R}^{3,3}$ represents the change of basis. Hence using the linearity of S we can write

$$\langle e_i, Se_j\rangle = \langle a_{1i}v_1 + a_{2i}v_2 + a_{3i}n, a_{1j}Sv_1 + a_{2j}Sv_2 + a_{3j}Sn\rangle = \sum_{k,l=1}^2 a_{ki}a_{lj}g(v_k, Sv_l) = \sum_{k,l=1}^2 a_{ki}a_{lj}h_{kl}\,.$$

An analogous computation for $\langle e_i, S_\phi^* e_j\rangle$ and the identity $(AA^T)_{i,j\le 2} = g^{-1}$ yield the result. □

Full elastic model and dissimilarity measure. Given a surface $\mathcal{S} \subset \mathbb{R}^3$ representing a physical shell with thickness $\delta > 0$ and a deformation $\phi : \mathcal{S} \to \mathbb{R}^3$, a generic elastic deformation energy is given by

$$\int_{\mathcal{S}} \delta\, W_{\text{mem}}(\mathcal{G}[\phi]) + \delta^3\, W_{\text{bend}}(S_\phi^{\text{rel}})\,\mathrm{d}a\,, \tag{2.22}$$

with $W_{\text{mem}}(A) = W_{\text{mem}}(\operatorname{tr} A, \det A)$ as defined in (2.3) for $d = 2$ and W_{bend} as defined in (2.19). Nevertheless, for convenience we shall consider in the following a rescaled version of (2.22), namely

$$\mathcal{W}_{\mathcal{S}}[\phi] = \int_{\mathcal{S}} W_{\text{mem}}(\mathcal{G}[\phi]) + \eta\, W_{\text{bend}}(S_\phi^{\text{rel}})\,\mathrm{d}a\,, \tag{2.23}$$

where the bending weight η represents the squared thickness of the shell. Note that (2.23) is invariant with respect to rigid body motions by construction, *i.e.* $\mathcal{W}_{\mathcal{S}}[\phi] = 0$ and $d\mathcal{W}_{\mathcal{S}}[\phi] = 0$ if $\phi(x) = Qx + b$ with $Q \in SO(3)$ and $b \in \mathbb{R}^3$. In particular, we have

$$\mathcal{W}_{\mathcal{S}}[\mathrm{id}] = 0, \quad d\mathcal{W}_{\mathcal{S}}[\mathrm{id}] = 0. \tag{2.24}$$

Similar tp (2.5), we can derive a *dissimilarity measure* for two given shells $\mathcal{S}_A, \mathcal{S}_B \subset \mathbb{R}^3$ by minimizing (2.23) over all deformations satisfying $\phi(\mathcal{S}_A) = \mathcal{S}_B$, *i.e.*

$$\mathrm{d}^2_{\text{shell}}(\mathcal{S}_A, \mathcal{S}_B) = \min_{\phi:\phi(\mathcal{S}_A)=\mathcal{S}_B} \int_{\mathcal{S}} W_{\text{mem}}(\mathcal{G}[\phi]) + \eta\, W_{\text{bend}}(S_\phi^{\text{rel}})\,\mathrm{d}a\,. \tag{2.25}$$

However, we do not discuss whether this definition is actually well-defined, *i.e*. if there exists such a minimizer. Physically, one might regard the second shell $\mathcal{S}_B$ as a deformed version of the first shell $\mathcal{S}_A$, *i.e*., the corresponding material of $\mathcal{S}_B$ is just in a deformed configuration compared to its configuration in $\mathcal{S}_A$. Since every material point has a well-defined position, one can view this correspondence as *a priori* information. In this setup, one can then assume the dissimilarity measure to be well-defined. However, we will see in the next section that the well-posedness of the corresponding discrete dissimilarity measure is trivial by construction of the discrete shell space.

Remark: Note the relation between the dissimilarity measure (2.25) and Thm. 1.7: $\mathrm{d}^2_{\text{shell}}(\mathcal{S}_A, \mathcal{S}_B) = 0$ iff. $\mathcal{S}_A$ and $\mathcal{S}_B$ are congruent, *i.e*. they differ only by a rigid body motion.

2.3 Discrete shells and discrete deformation energies

According to Def. 1.13, a *discrete surface* is a triangular mesh $\mathcal{M}_h = (\mathcal{V}, \mathcal{F})$ along with an injective embedding $E : \mathcal{V} \to \mathbb{R}^3$, such that $\mathcal{M}_h$ is a discrete 2-manifold which is orientable. In particular, a discrete surface is uniquely determined by its geometry and connectivity. We assume that $\mathcal{M}_h$ is a polyhedral approximation of the midsurface $\mathcal{S}$ of a thin elastic shell. Hence we will denote a discrete surface $\mathbf{S} := \mathcal{M}_h$ that has also this physical interpretation as a *discrete shell* - however, these terms are interchangeable.

In the following we consider (discrete) deformations between different discrete surfaces (resp. discrete shells). Alternatively, we assume a designated discrete *reference surface* to be given which prescribes the connectivity/topology encoded in the sets $\mathcal{V}$ and $\mathcal{F}$. Then different discrete shells are given by different embeddings of the topologically identical mesh.

Definition 2.5 (Dense correspondence and discrete deformation)**.** We say that two discrete surfaces (resp. discrete shells) are in *dense correspondence* or in 1-to-1-correspondence if they share the same connectivity/topology. Given two discrete surfaces $\mathbf{S}$ and $\tilde{\mathbf{S}}$ with embeddings $E : \mathcal{V} \to \mathbb{R}^3$ and $\tilde{E} : \mathcal{V} \to \mathbb{R}^3$, respectively, which are in dense correspondence. A discrete deformation $\Phi : \mathbf{S} \to \tilde{\mathbf{S}}$ is the unique piecewise affine mapping defined by its nodal values $\Phi(E(v_i)) := \tilde{E}(v_i)$ for $i = 1, \dots, |\mathcal{V}|$.

That means, if $X = (1-\xi_1-\xi_2)X_i + \xi_1 X_j + \xi_2 X_k \in T(f)$, where $T(f) = \{X_i, X_j, X_k\} \subset \mathbf{S}$ and $\xi_1, \xi_2 \in [0,1]$ barycentric coordinates, we have $\Phi(X) = (1-\xi_1-\xi_2)\tilde{X}_i + \xi_1 \tilde{X}_j + \xi_2 \tilde{X}_k$ with $\tilde{T}(f) = \{\tilde{X}_i, \tilde{X}_j, \tilde{X}_k\} \subset \tilde{\mathbf{S}}$.

Remark: The (pairwise) dense correspondence will guarantee the well-definedness of a dissimilarity measure on the space of discrete surfaces (*cf*. eq. (2.25)).

In the following we will consider families of discrete surfaces which are pairwise in dense correspondence. Actually, dense correspondence defines an equivalence relation, *i.e*. we consider a fixed equivalence class. This means, all discrete surfaces are based on the same sets of indices $\mathcal{V}$ and $\mathcal{F}$, respectively. Nevertheless, for two discrete surfaces $\mathbf{S}$ and $\tilde{\mathbf{S}}$ we will often denote the corresponding sets by $\mathcal{V}, \mathcal{F}$ and $\tilde{\mathcal{V}}, \tilde{\mathcal{F}}$, respectively. This convention simp[lifies the notation, *e.g*. a_f refers to the area of face f in $\mathbf{S}$, whereas $a_{\tilde{f}} := \tilde{a}_f$ denotes the area of face f in $\tilde{\mathbf{S}}$. Here we have to keep in mind that for $f_i \in \mathcal{F}$ and $\tilde{f}_i \in \tilde{\mathcal{F}}$ we actually have $f_i = \tilde{f}_i$ but $T(f_i) \neq \tilde{T}(\tilde{f}_i)$, since $\mathbf{S}$ and $\tilde{\mathbf{S}}$ have different embeddings.

Discrete membrane model. Let $\mathbf{S}$ and $\tilde{\mathbf{S}}$ two discrete surfaces resp. discrete shells that are in dense correspondence. We have elementwise constant first fundamental forms and a unique correspondence between all faces. Furthermore, we want to make use of the membrane model derived in Sec. 2.2 and in particular of the representation of the distortion tensor (2.13). Hence to describe membrane distortions induced by a discrete deformation $\Phi : \mathbf{S} \to \tilde{\mathbf{S}}$ we arrive at an elementwise constant, discrete distortion tensor

$$\mathcal{G}[\Phi]|_f = (G_f)^{-1} G_f^\Phi \in \mathbb{R}^{2,2}, \quad f \in \mathcal{F}. \tag{2.26}$$

Here and in the following all quantities with either a tilde or a super index Φ are living on the deformed surface $\tilde{\mathbf{S}} = \Phi(\mathbf{S})$, *e.g*. if $G_f \in \mathbb{R}^{2,2}$ is the discrete first fundamental form on the face $f \in \mathcal{F}$ in the *undeformed* configuration $\mathbf{S}$, then $\tilde{G}_f = G_f^\Phi \in \mathbb{R}^{2,2}$ is the corresponding form on the *deformed* configuration $\tilde{\mathbf{S}} = \Phi(\mathbf{S})$.

Using the continuous membrane model in Sec. 2.2 and in particular the generic membrane energy in (2.14) we define the *discrete membrane energy* by

$$\mathbf{W}_{\text{mem}}[\mathbf{S}, \tilde{\mathbf{S}}] = \int_{\mathbf{S}} W_{\text{mem}}(\mathcal{G}[\Phi])\,\mathrm{d}a = \sum_{f\in\mathcal{F}} a_f \cdot W_{\text{mem}}(\mathcal{G}[\Phi]|_f)\,, \quad \tilde{\mathbf{S}} = \Phi(\mathbf{S})\,. \tag{2.27}$$

Note that a one point quadrature is sufficient as we are dealing with an elementwise constant discrete distortion tensor. Different from (2.14), the discrete membrane energy directly depends on the undeformed and deformed discrete shell. For the membrane energy density W_{mem} we can use exactly the same density as in (2.14), *i.e.*

$$W_{\text{mem}}(\mathcal{G}[\Phi]|_f) = \frac{\mu}{2}\mathrm{tr}\,\mathcal{G}[\Phi]|_f + \frac{\lambda}{4}\det\mathcal{G}[\Phi]|_f - \left(\frac{\mu}{2}+\frac{\lambda}{4}\right)\log\det\mathcal{G}[\Phi]|_f - \mu - \frac{\lambda}{4}\,.$$

Note that $\mathrm{tr}\,\mathcal{G}[\Phi]|_f$ controls the local change of length, *i.e.* the change of edge lengths, whereas $\det\mathcal{G}[\Phi]|_f$ controls the local change of volume, *i.e.* the change of triangle volumes. In particular, the density grows quadratically for $\det\mathcal{G}[\Phi]|_f \to \infty$ but due to the log-term it grows even faster for $\det\mathcal{G}[\Phi]|_f \to 0$. This prevents a local interpenetration of matter, *i.e.* the degeneration of triangles. Finally, we have $W_{\text{mem}}(\mathbb{1}) = 0$ and $dW_{\text{mem}}(\mathbb{1}) = 0$.

Discrete bending model. Having a notion of a discrete shape operator given by (1.23) at hand, we can translate the general representation of a bending energy given in (2.18) (with the density (2.19) and $\alpha \in \{0, 1\}$) directly into the discrete setup. Setting $\alpha = 0$, as in (2.20), we can define a discrete bending energy via

$$\mathbf{W}_{\text{bend}}[\mathbf{S}, \tilde{\mathbf{S}}] = \sum_{f\in\mathcal{F}} a_f \cdot \mathrm{tr}\left((S_f - S_f^{\Phi})^2\right), \quad \tilde{\mathbf{S}} = \Phi(\mathbf{S})\,, \tag{2.28}$$

with $a_f = |T(f)|$ as above. Alternatively, by choosing $\alpha = 1$, as in (2.21), we can derive a discrete version of the Willmore energy:

$$\tilde{\mathbf{W}}_{\text{bend}}[\mathbf{S}, \tilde{\mathbf{S}}] = \sum_{f\in\mathcal{F}} a_f \cdot \left(\mathrm{tr}\,(S_f - S_f^{\Phi})\right)^2, \quad \tilde{\mathbf{S}} = \Phi(\mathbf{S})\,. \tag{2.29}$$

Note that a one point quadrature is again sufficient as we are integrating over an elementwise constant density.

A simplified discrete bending model. In the remainder of this section we investigate another definition of a discrete Willmore energy and derive a representation that corresponds to a non-conforming FEM approach. Furthermore, after some simplifications, we obtain the *Discrete Shells* bending model [GHDS03] as a special case.

In (1.24), we have computed a triangle-averaged mean curvature, *i.e.*

$$\mathrm{tr}\,S_f = \mathrm{tr}\,(G_f^{-1}H_f) = -\sum_{i=0}^{2}\frac{\cos\frac{\theta_i}{2}}{a_f}\|E_i\|\,.$$

Now the discrete mean curvature functional can be written as a sum over edges:

$$\int_{\mathbf{S}} \mathrm{tr}\,S\,\mathrm{d}a = \sum_{f\in\mathcal{F}} a_f \cdot \mathrm{tr}\,S_f = \sum_{f\in\mathcal{F}}\sum_{i=0}^{2} -\cos\frac{\theta_i}{2}\,\|E_i\| = \sum_{e\in\mathcal{E}} -2\,\cos\frac{\theta_e}{2}\,l_e\,,$$

with l_e being the length of edge $e \in \mathcal{E}$. We introduce an area $d_e = \frac{1}{3}(a_f + a_{f'})$ if e is an edge of the two adjacent faces f and f', *cf.* Fig. 3. Then we rewrite the discrete mean curvature functional by introducing a mean curvature density at edges:

$$\int_{\mathbf{S}} \mathrm{tr}\,S\,\mathrm{d}a = \sum_{e\in\mathcal{E}} d_e \cdot \left(\frac{-2\,\cos\frac{\theta_e}{2}}{d_e}\,l_e\right). \tag{2.30}$$

Now we further simplify (2.30) to derive the *Discrete Shells* bending model proposed in [GHDS03]. A Taylor expansion of the function $f(\theta) = -2\,\cos\frac{\theta}{2}$ about $\theta = \pi$ yields $f(\theta) = (\theta - \pi) + O(|\theta - \pi|^3)$. Let $\mathbf{S}$ be a reference

shell, *i.e.* the undeformed configuration. If we assume that we have an isometric[3] deformation $\Phi : \mathbf{S} \to \mathbb{R}^3$, *i.e.* $l_e^\Phi = l_e$ and $d_e^\Phi = d_e$, we obtain up to higher order terms

$$\int_{\mathbf{S}} \operatorname{tr}\left(S - S_\Phi \circ \Phi\right) \mathrm{d}a = \sum_{e \in \mathcal{E}} d_e \left(\frac{\theta_e - \theta_e^\Phi}{d_e} \, l_e \right) .$$

Here $\theta_e^\Phi = \tilde{\theta}_e$ is the dihedral angle at edge e in the deformed mesh $\tilde{\mathbf{S}} = \Phi(\mathbf{S})$. In the spirit of [MDSB02], one arrives at the *Discrete Shells* bending model by squaring the discrete density, *i.e.*

$$\mathbf{W}_{\text{bend}}^{\text{DS}}[\mathbf{S}, \tilde{\mathbf{S}}] = \sum_{e \in \mathcal{E}} \frac{(\theta_e - \tilde{\theta}_e)^2}{d_e} \, l_e^2 \,, \quad \tilde{\mathbf{S}} = \Phi(\mathbf{S}) \,. \tag{2.31}$$

Intuitively, $\mathbf{W}_{\text{bend}}^{\text{DS}}$ can be considered as a simplification of (2.29). Although (2.31) coincides exactly with the *Discrete Shells* bending energy introduced in [GHDS03], the authors in [GHDS03] derive their discrete bending energy by using results from [CSM03].

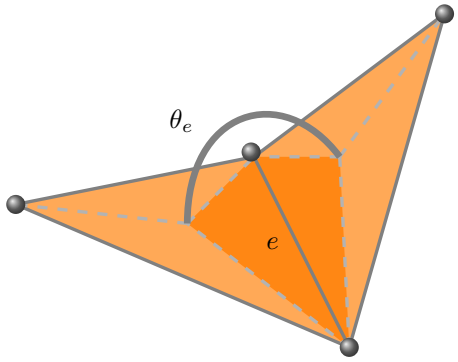


Figure 3: Support of the *Discrete Shells* bending energy [GHDS03]; the dihedral angle $\theta_e = \alpha_e - \pi$ at an edge e is defined as the angle between adjacent triangle normals, where α_e is the angle between the two faces. The darker region represents the area d_e associated with e.

A discrete dissimilarity measure. Finally, we are able to define a dissimilarity measure on the space of discrete surfaces and discrete shells, respectively, given by a discrete deformation energy:

Definition 2.6 (Discrete dissimilarity measure)**.** Given two discrete surfaces $\mathbf{S} = (\mathcal{N}, \mathcal{F}, \mathcal{E})$ and $\tilde{\mathbf{S}} = (\tilde{\mathcal{N}}, \tilde{\mathcal{F}}, \tilde{\mathcal{E}})$ that are in dense correspondence, *i.e.* there is a unique affine deformation Φ with $\tilde{\mathbf{S}} = \Phi(\mathbf{S})$. The *discrete deformation energy* $\mathbf{W} = \mathbf{W}_{\mathbf{S}}[\Phi] = \mathbf{W}[\mathbf{S}, \tilde{\mathbf{S}}]$ is defined by

$$\mathbf{W}[\mathbf{S}, \tilde{\mathbf{S}}] = \mathbf{W}_{\text{mem}}[\mathbf{S}, \tilde{\mathbf{S}}] + \eta \, \mathbf{W}_{\text{bend}}[\mathbf{S}, \tilde{\mathbf{S}}] \,,$$

where the *bending weight* $\eta = \delta^2$ represents the squared thickness of the shell. The *discrete membrane energy* and the *discrete bending energy*, respectively, are given by

$$\mathbf{W}_{\text{mem}}[\mathbf{S}, \tilde{\mathbf{S}}] = \sum_{f \in \mathcal{F}} a_f \cdot W_{\text{mem}}(\mathcal{G}[\Phi]|_f) \,,$$
$$\mathbf{W}_{\text{bend}}[\mathbf{S}, \tilde{\mathbf{S}}] = \sum_{f \in \mathcal{F}} a_f \cdot W_{\text{bend}}(S_f - S_f^\Phi) \,,$$

where $\mathcal{G}[\Phi] \in \mathbb{R}^{2,2}$ denotes the discrete distortion tensor defined in (2.26) and $S \in \mathbb{R}^{2,2}$ the matrix representation of the discrete shape operator defined in (1.23). The membrane density W_{mem} and the bending density W_{bend},

[3] When deriving (discrete) bending models, one typically assumes to deal with *inextensible* materials which are characterized by mostly isometric deformations, *cf*. *e.g.* [GHDS03, BWH+06]. This corresponds to the analytic results presented in Sec. 2.2, where bending modes are of higher order and hence only decisive when the present deformation is (almost) isometric.

respectively, are defined as

$$W_{\text{mem}}(A) = \frac{\mu}{2}\operatorname{tr} A + \frac{\lambda}{4}\det A - \left(\mu + \frac{\lambda}{2}\right)\log\det A - \mu - \frac{\lambda}{4}\,,$$
$$W_{\text{bend}}(A) = \alpha(\operatorname{tr} A)^2 + (1-\alpha)\operatorname{tr}(A^2)\,, \quad \alpha \in \{0,1\}\,.$$

For $\alpha = 1$, a simplification leads to the *Discrete Shells* bending energy [GHDS03], *i.e.*

$$\mathbf{W}_{\text{bend}}^{\text{DS}}[\mathbf{S}, \tilde{\mathbf{S}}] = \sum_{e\in\mathcal{E}} \frac{(\theta_e - \tilde{\theta}_e)^2}{d_e}\, l_e^2\,.$$

3 The shape space of discrete surfaces

It is due to Kendall [Ken84] that complex shapes, *e.g.* curves, images or solid materials, are considered as individual elements or points in a high or even infinite dimensional space, *i.e.* the *shape space*. Initially, this space is just a collection of shapes without any mathematical structure. In particular, most shape spaces cannot be considered as linear vector spaces. Nevertheless, one is interested in performing mathematical operations on the set of shapes, for instance, for two given shapes one wants to compute a connecting path (*cf.* Fig. 4). The notion of an optimal or shortest path then induces naturally a distance measure which allows *e.g.* for a statistical analysis. It is a well-established ansatz to consider a given shape space as a *Riemannian manifold*. In a nutshell, a Riemannian manifold can be described as a collection of points that is locally equivalent to the Euclidean space, together with a so-called *Riemannian metric*, *i.e.* an instruction how to measure local variations. On a Riemannian manifold the notion of a connecting path and hence a (locally) shortest path, a so-called *geodesic*, is intrinsically given. Thereby, a geodesic connecting two points can be considered as the solution of the interpolation problem. Similarly, one can extrapolate by extending geodesic paths via the *exponential map* or transport details via the *parallel transport*—both are inherent concepts in Riemannian manifold theory. Hence the mathematical structure of a Riemannian manifold leads to the solution of a couple of problems relevant *e.g.* in computer graphics.

Figure 4: Morphing by means of interpolation (orange) computed between two input shapes (gray), *cf.* [HRWW12].

In this section we shall consider the shape space of discrete surfaces as a *Riemannian manifold*. We aim at combining a physically sound model of thin shells (as it has been derived in Sec. 2) with a consistent definition of geometric objects in a Riemannian manifold. The key ingredient, as we will see, is the computation of (locally) shortest paths in the manifold, *i.e. geodesic curves*. Hence the collection of geometric objects and corresponding operators in the Riemannian framework is also referred to as *geodesic calculus*.

Continuous geodesics are minimizers of the so-called path energy. One way to approximate geodesic paths connecting two points in a generic manifold is via the minimization of a discretized path energy. Instead of discretizing the underlying flow, the *variational time-discretization* proposed by Rumpf and Wirth [WBRS11] is based on the direct minimization of this discrete path energy subject to the prescribed data given at the initial and the end time. In particular, this approach is built on a local approximation of the squared Riemannian distance, where this approximation can be thought of as a *dissimilarity measure* between shapes. Hence, we shall use the (discrete) dissimilarity measure derived in the previous section to apply the variational time-discretization to the space of discrete shells in order to compute (time-discrete) geodesics.
Building on the variational time-discretization of geodesic paths, Rumpf and Wirth [RW13] developed a comprehensive discrete geodesic calculus on the space of viscous fluidic objects and presented in [RW15] a corresponding complete convergence analysis on general finite- and on certain infinite-dimensional shape spaces with the structure of a Banach manifold. The generic definitions of several discrete geometric objects, such as exponential map, logarithm and parallel transport, are appropiate to be transferred directly to other shape spaces. To this end, we apply exactly this discrete geodesic calculus to the space of discrete shells to obtain useful and robust tools for applications in computer graphics such as extrapolation or detail transfer.

3.1 Geodesic calculus on a Riemannian manifold

Geodesics are usually defined as curves that transport their velocity vector parallely or equivalently, that are solutions of the *geodesic equation*. Both definitions are based on the notion of a *covariant derivative* along a curve. However, we first define geodesics as minimizers of the *path energy*. Later, we will see that this is well-defined as these minimizers indeed satisfy the geodesic equation.

References: The essential results discussed in this section have been presented by Martin Rumpf and Benedikt Wirth in [RW15]. All concepts based on finite dimensional Riemannian geometry are presented according to the well-established textbook by M. doCarmo [dC92]; for further reading on infinite dimensional manifolds we refer to the textbook by S. Lang [Lan95].

We define a *differentiable manifold* $\mathcal{M}$ of dimension $d < \infty$ in the sense of Definition 2.1 in [dC92, chap. 0], *i.e.* there is a family of injective mappings $x_\alpha : \omega_\alpha \subset \mathbb{R}^d \to \mathcal{M}$ with $\cup_\alpha x_\alpha(\omega_\alpha) = \mathcal{M}$, such that $x_\beta^{-1} \circ x_\alpha$ is differentiable for any pair α, β with $x_\alpha(\omega_\alpha) \cap x_\beta(\omega_\beta) \neq \varnothing$. For convenience, we will assume in the following that there is one global parametrization $x : \omega \subset \mathbb{R}^d \to \mathcal{M}$ with $x(\omega) = \mathcal{M}$. In particular, x is twice differentiable, injective and regular in the sense that Dx has full rank. The tangent space $T_p\mathcal{M}$ of $\mathcal{M}$ at $p \in \mathcal{M}$ is defined as

$$T_p\mathcal{M} = \{\dot\gamma(0) \mid \gamma : (-\epsilon, \epsilon) \to \mathcal{M} \text{ is a smooth curve with } \gamma(0) = p,\ \epsilon > 0\}\,.$$

If $x : \omega \to \mathcal{M}$ is a parametrization with $x(\xi) = p$ for some $\xi = (\xi_1, \dots, \xi_d) \in \omega$, the set $(X_1, \dots, X_d)$ with $X_i = X_i(p) = X_i(\xi) = x_{,i}(\xi) = \frac{\partial x}{\partial \xi_i}(\xi)$ is a basis of $T_p\mathcal{M}$, denoted as *canonical basis*. A vector field V on $\mathcal{M}$ is a mapping with $V(p) \in T_p\mathcal{M}$ for all $p \in \mathcal{M}$.

A *Riemannian metric* on $\mathcal{M}$ is a mapping $g : p \mapsto g_p$ such that $g_p : T_p\mathcal{M} \times T_p\mathcal{M} \to \mathbb{R}$ is a bilinear, symmetric and positive-definite form, which varies smoothly in the sense that $\xi \mapsto g_{ij}(\xi) := g_{x(\xi)}(X_i(\xi), X_j(\xi))$ is a differentiable function in ω. A manifold equipped with a Riemannian metric is referred to as Riemannian manifold. As $(g_{ij})_{ij}$ is a regular matrix in $\mathbb{R}^{d,d}$ there is an inverse matrix $g^{-1} \in \mathbb{R}^{d,d}$ which is denoted by $(g^{kl})_{kl}$, *i.e.* $g_{ij}g^{jk} = \delta_{ik}$.

Remark: To avoid confusion, we denote the metric associated with generic Riemannian manifolds by *Riemannian metric* and the metric on a two-dimensional embedded surface by *first fundamental form*, *cf*. Sec. 1.2.

Path energy and geodesics. Given a smooth path $(y(t))_{t\in[0,1]}$ on a Riemannian manifold $(\mathcal{M}, g)$, the length of this path is defined as

$$\mathcal{L}[(y(t))_{t\in[0,1]}] = \int_0^1 \sqrt{g_{y(t)}(\dot y(t), \dot y(t))}\,\mathrm{d}t\,. \tag{3.1}$$

Note that the path length is independent of reparameterization. This geometrically nice property leads to analytical complications when dealing with the existence theory of shortest paths as well as to computational difficulties when optimizing this non-convex functional. The path energy is defined as

$$\mathcal{E}[(y(t))_{t\in[0,1]}] = \int_0^1 g_{y(t)}(\dot y(t), \dot y(t))\,\mathrm{d}t\,. \tag{3.2}$$

In contrast to $\mathcal{L}$, the path energy is *not* independent of reparameterization. A direct application of the Cauchy-Schwarz inequality shows that

$$\mathcal{L}[(y(t))_{t\in[0,1]}] \le \sqrt{\mathcal{E}[(y(t))_{t\in[0,1]}]}$$

and equality holds if and only if $g_{y(t)}(\dot y(t), \dot y(t)) = \text{const}$. We will see that minimizers of $\mathcal{E}$ will have this *constant speed property*. Thus, to identify shortest paths for fixed boundary data $y(0) = y_A$ and $y(1) = y_B$ with $y_A,\ y_B \in \mathcal{M}$ we will seek for minimizers of the path energy and minimizers will be paths with constant speed.

Definition 3.1 (Geodesic path)**.** For $y_A,\ y_B \in \mathcal{M}$ a minimizer of the path energy among all path $y : [0, 1] \to \mathcal{M}$ with $y(0) = y_A$ and $y(1) = y_B$ is denoted as geodesic path connecting y_A and y_B.

Rumpf and Wirth have shown in [RW15] that this variational definition is indeed well-defined, *i.e.* a minimizer of $\mathcal{E}$ exists and is unique under suitable assumptions. In particular, their results holds for general, possibly infinite dimensional manifolds. We will discuss these results in the following.

Existence and uniqueness of geodesic paths. In this paragraph we present the theoretical existence and uniqueness results by Rumpf and Wirth [RW15]. In particular, the Riemannian manifold $\mathcal{M}$ does not necessarily be finite dimensional. To this end, we need a precise technical setup. Let $\mathbf{V}$ be a separable, reflexive Banach space that is compactly embedded[4] in a Banach space $\mathbf{Y}$. Let $\mathcal{M}$ be the closure of an open connected subset of $\mathbf{V}$ and hence a Banach manifold, potentially with boundary (in which case we assume the boundary $\partial\mathcal{M}$ to be smooth). We assume $\mathcal{M}$ to be path-connected. For some reference point $\hat{y} \in \mathbf{Y}$ we might think of $\mathcal{M}$ as being some affine space attached to $\hat{y}$, *i.e.* $\mathcal{M} = \hat{y} + \mathbf{V}$. In particular, the tangent bundle of $\mathcal{M}$ is associated with $\mathbf{V}$, *i.e.* $T_y\mathcal{M} = \mathbf{V}$ for all $y \in \mathcal{M}$. Let $g : \mathbf{Y} \times \mathbf{Y} \times \mathbf{Y} \to \mathbb{R}$ be a Riemannian metric with $g_y(v, v) = \infty$ if $v \notin \mathbf{V}$, which satisfies the following hypotheses:

$$\text{(H1)} \qquad \begin{cases} g \text{ is uniformly bounded and } \mathbf{V}\text{-coercive in the sense } c^*\|v\|_{\mathbf{V}}^2 \le g_y(v,v) \le C^*\|v\|_{\mathbf{V}}^2 . \\ g \text{ is continuous in the sense } |g_y(v,v) - g_{\tilde{y}}(v,v)| \le \beta(\|y-\tilde{y}\|_{\mathbf{Y}})\|v\|_{\mathbf{V}}^2 \\ \text{for a strictly increasing, continuous function } \beta \text{ with } \beta(0) = 0. \end{cases}$$

Hypothesis (H1) is globally fulfilled only for quite special Riemannian manifolds. However, the setup is also adequate to analyze general, possibly infinite-dimensional manifolds locally, where the linear space $\mathbf{V}$ or its subset $\mathcal{M}$ have to be interpreted as a chart of the considered manifold.

For $y_A,\ y_B \in \mathcal{M}$, the next theorem states the existence of a connecting path with least energy. The key point in the proof is the weak lower semi-continuity of the continuous path energy (3.2) using the compact embedding of $\mathbf{V}$ into $\mathbf{Y}$.

Theorem 3.2 (Existence of continuous geodesics, [RW15])**.** *Let $(\mathcal{M}, g)$ be a Riemannian manifold satisfying assumption* (H1)*. Then the energy* (3.2) *is lower semi-continuous with respect to weak convergence in $H^1((0,1);\mathcal{M})$. Furthermore, for $y_A,\ y_B \in \mathcal{M}$ there exists a geodesic connecting y_A and y_B,* i.e. *a minimizer of $\mathcal{E}$ in the space of all paths $(y(t))_{t\in[0,1]} \in H^1((0,1);\mathcal{M})$ with $y(0) = y_A$ and $y(1) = y_B$. In particular, y is Hölder continuous (in the $\mathbf{V}$-topology).*

Proof: see Thm 4.1 in [RW15]. □

As in finite-dimensional Riemannian geometry the shortest geodesic between close points is unique as stated in the next theorem (*cf*. also Cor. 5.2 in [Lan95, VIII]).

Theorem 3.3 (Uniqueness of short continuous geodesics, [RW15])**.** *Under the assumptions of Theorem 3.2, for the metric g being $C^2(\mathcal{M}; \mathbf{V}' \otimes \mathbf{V}')$-smooth, geodesics are unique locally. That means, given $y_A \in \mathcal{M}$ there is some small $\delta > 0$ depending on y_A such that for each $y_B \in \mathcal{M}$ with $\|y_A - y_B\|_{\mathbf{V}} < \delta$ the shortest geodesic between y_A and y_B is unique.*

Proof: see Thm 4.2 in [RW15]. □

Once we have existence and uniqueness of geodesics, we can define a Riemannian distance of two points $y_A,\ y_B \in \mathcal{M}$ in the usual way, *i.e.*

$$\operatorname{dist}(y_A, y_B) = \min_{y(0)=y_A, y(1)=y_B} \mathcal{L}[(y(t))_{t\in[0,1]}] = \sqrt{\min_{y(0)=y_A, y(1)=y_B} \mathcal{E}[(y(t))_{t\in[0,1]}]}\,. \tag{3.3}$$

One can verify the axioms of a metric and show that the induced topology is equivalent to the $\mathbf{V}$-topology, *i.e.* $\sqrt{c^*}\|y_B - y_A\|_{\mathbf{V}} \le \operatorname{dist}(y_A, y_B) \le \sqrt{C^*}\|y_B - y_A\|_{\mathbf{V}}$.

Covariant derivative. Next, we will investigate the differentiation of vector fields on manifolds, which will lead us to the notion of the covariant derivative. Finally, we will explore a connection to the Euler-Lagrange equations of geodesic paths. Here we will focus on the covariant derivative of a vector field *along a curve* $y : [0,1] \to \mathcal{M}$. In practice, most important is the vector field given by the curve's velocity field $v(t) := \dot{y}(t) \in T_{y(t)}\mathcal{M}$, $t \in [0,1]$. However, $\dot{v}(t) = \ddot{y}(t)$ is not a tangent vector in general. The covariant derivative $\frac{D}{dt}$ resolves this "problem" since $\frac{D}{dt}v(t) \in T_{y(t)}\mathcal{M}$ again.

[4] That means, $\mathrm{id} : \mathbf{V} \to \mathbf{Y}$ is a bounded and compact operator, i.e. $\|v\|_{\mathbf{Y}} \le c\|v\|_{\mathbf{V}}$ for some $c > 0$ and all $v \in \mathbf{V}$ and each bounded subsequence in $\|.\|_{\mathbf{V}}$ has a converging subsequence wrt. $\|.\|_{\mathbf{Y}}$.

We first consider a finite dimensional Riemannian manifold $\mathcal{M}$ with a parametrization $x : \Omega \subset \mathbb{R}^d \to \mathcal{M}$. This allows us to work with *coordinates*, that means, we can express all quantities in a finite dimensional basis of the tangent space. However, to transfer the following concepts to the infinite dimensional setup, we eventually aim at deriving a *coordinate-free* representation.

The second derivatives $x_{,ij} = \partial_{\xi_i}\partial_{\xi_j} x$ of the parametrization $x : \Omega \subset \mathbb{R}^d \to \mathcal{M}$ decomposes into a normal component and into a tangential component, *i.e.* we obtain

$$x_{,ij} = \sum_{k=1}^{d} \Gamma_{ij}^k X_k + \sum_l \beta_l n_l \,,$$

where Γ_{ij}^k are the Christoffel symbols and $(n_l)_l$ is a basis of the normal space $(T_p\mathcal{M})^\perp$. It follows that

$$x_{,ij} \cdot X_k = \sum_{l=1}^{d} \Gamma_{ij}^l X_l \cdot X_k = \sum_{l=1}^{d} \Gamma_{ij}^l g_{lk} \,. \tag{3.4}$$

From the symmetry of second derivative of the parametrization we immediately obtain the symmetry of the Christoffel symbols, *i.e.* $\Gamma_{ij}^k = \Gamma_{ji}^k$.

Proposition 3.4. *Let $g^{-1} = (g^{ij})_{ij=1,\dots,d}$ be the inverse of g. Then we obtain the following representation of the Christoffel symbols:*

$$\Gamma_{ij}^k = \frac{1}{2} \sum_{l=1}^{d} g^{lk} \left(g_{jl,i} - g_{ij,l} + g_{li,j} \right) . \tag{3.5}$$

Proof: Differentiation of the metric and taking into account (3.4) gives

$$\begin{aligned}
g_{jl,i} &= x_{,ji} \cdot x_{,l} + x_{,j} \cdot x_{,li} = \sum_{m=1}^{d} \Gamma_{ji}^m g_{ml} + \sum_{m=1}^{d} \Gamma_{li}^m g_{mj} \,, \\
g_{ij,l} &= x_{,il} \cdot x_{,j} + x_{,i} \cdot x_{,jl} = \sum_{m=1}^{d} \Gamma_{il}^m g_{mj} + \sum_{m=1}^{d} \Gamma_{jl}^m g_{mi} \,, \\
g_{li,j} &= x_{,lj} \cdot x_{,i} + x_{,l} \cdot x_{,ij} = \sum_{m=1}^{d} \Gamma_{lj}^m g_{mi} + \sum_{m=1}^{d} \Gamma_{ij}^m g_{ml} \,.
\end{aligned}$$

Summing the first and the third equation and substracting the second equation we obtain

$$g_{jl,i} - g_{ij,l} + g_{li,j} = 2 \sum_{m=1}^{d} \Gamma_{ij}^m g_{ml} \,,$$

which after multiplication with g^{-1} (*i.e.* applying $\sum_l g^{lk}(\dots)$ on both sides) verifies the claimed representation. □

To obtain a *coordinate-free* formulation, we define a bilinear operator $\Gamma = \Gamma_p : T_p\mathcal{M} \times T_p\mathcal{M} \to T_p\mathcal{M}$ by

$$\Gamma(X_i, X_j) = \sum_{k=1}^{d} \Gamma_{ij}^k X_k \,,$$

and get for tangent vectors $U = \sum_i u_i X_i$, $V = \sum_j v_j X_j$ and $W = \sum_l w_l X_l$:

$$g_p(\Gamma(U,V), W) = \sum_{i,j,k,l=1}^{d} u_i \, v_j \, w_l \, \Gamma_{ij}^l g_{kl} \,.$$

On the other hand, testing the right-hand side of (3.5) with U, V, W in the metric yields

$$g_p(\Gamma(U,V), W) = \frac{1}{2} \Big(\left(D_p g\right)(V)(U,W) + \left(D_p g\right)(U)(V,W) - \left(D_p g\right)(W)(U,V) \Big) \,.$$

Definition 3.5 (Christoffel operator)**.** For $p \in \mathcal{M}$ the Christoffel operator $\Gamma = \Gamma_p$ is a mapping $\Gamma_p : T_p\mathcal{M} \times T_p\mathcal{M} \to T_p\mathcal{M}$. For $U, V \in T_p\mathcal{M}$ the evaluation $\Gamma_p(U, V)$ is defined implicitly by

$$g_p(\Gamma_p(U,V),W) = \frac{1}{2}\Big((D_p g)(V)(U,W) + (D_p g)(U)(V,W) - (D_p g)(W)(U,V)\Big) \quad \forall W \in T_p\mathcal{M}\,.$$

Remark: Existence of such an operator is shown in Thm. 4.2 of [Lan95, VIII].

Let $I \subset \mathbb{R}$ be a compact interval, e.g. $I = [0,1]$, and $y : I \to \mathcal{M}$ be a curve and W a vector field along the curve, *i.e.* $W(t) = \sum_{l=1}^d w_l(t) X_l(y(t))$. Let $\dot W(t) := \sum_{l=1}^d \dot w_l(t) X_l(y(t))$ and let $V(t) = \dot y(t)$ with $V(t) = \sum_l v_l(t) X_l(y(t))$, which is also a vector field along y. The product rule implies

$$\frac{d}{dt}W(t) = \sum_{l=1}^d \left(\dot w_l(t) X_l(y(t)) + w_l(t) \sum_{k=1}^d x_{,lk}(y(t)) v_k(t) \right)\,.$$

The covariant derivative $\frac{D}{dt}W(t)$ of W along y is defined as the projection of $\frac{d}{dt}W(t)$ onto the tangent space. Hence we replace $x_{,lk}$ by its tangential part, *i.e.* $\sum_{m=1}^d \Gamma_{lk}^m X_m = \Gamma_{y(t)}(X_l, X_k)$, which leads to the following definition which is also *coordinate-free*:

Definition 3.6 (Covariant derivative)**.** Let $y : I \to \mathcal{M}$ be a curve and $W : I \to T\mathcal{M}$ a vector field along y. If $\dot y(t) = \sum_l v_l(t) X_l(y(t))$, we define the covariant derivative $\frac{D}{dt}W$ of W along y for $t \in I$ by

$$\frac{D}{dt}W(t) = \sum_{l=1}^d \left(\dot w_l(t) X_l(y(t)) + w_l(t) \sum_{k=1}^d \Gamma_{y(t)}(X_l, X_k) v_k(t) \right) = \dot W(t) + \Gamma_{y(t)}(W(t), \dot y(t))\,. \tag{3.6}$$

Remark: Due to its coordinate-free formulation, Def. 3.6 is also valid for infinite dimensional manifolds (*cf.* [Lan95]). Since the following concepts are based on this definition they are not restricted to finite dimensional manifolds, either.

As mentioned before, a curve $y : I \to \mathcal{M}$ is usually defined to be *geodesic* if it solves the *geodesic equation*, *i.e.*

$$\frac{D}{dt}\dot y(t) = 0 \quad \forall t \in I\,.$$

The next theorem states that geodesics as defined in Def. 3.1 are solutions of the geodesic equation:

Theorem 3.7. *If $y : [0,1] \to \mathcal{M}$ is a geodesic connecting $y(0)$ and $y(1)$, then $\frac{D}{dt}\dot y(t) = 0$ for all $t \in (0,1)$.*

Proof. Consider the Euler–Lagrange equation of the path energy and apply integration by parts to obtain

$$\begin{aligned} 0 &= \partial_y \mathcal{E}[y](\vartheta) = \int_0^1 (D_y g_y)(\vartheta)(\dot y, \dot y) + 2 g_y(\dot y, \dot\vartheta)\,\mathrm{d}t \\ &= \int_0^1 (D_y g_y)(\vartheta)(\dot y, \dot y) - 2(D_y g_y)(\dot y)(\dot y, \vartheta) - 2 g_y(\ddot y, \vartheta)\,\mathrm{d}t \end{aligned}$$

for all smooth test vector fields ϑ along the path y. By the fundamental lemma we achieve

$$0 = g_y(\ddot y, \vartheta) + (D_y g_y)(\dot y)(\dot y, \vartheta) - \frac{1}{2}(D_y g_y)(\vartheta)(\dot y, \dot y) = g_y(\ddot y + \Gamma(\dot y, \dot y), \vartheta) = g_y(\frac{D}{\mathrm{d}t}\dot y, \vartheta)\,. \quad \square$$

In particular, minimizers y of the path energy satisfy a *constant speed property*, *i.e.* $g_{y(t)}(\dot y(t), \dot y(t)) = const.$

3.2 Variational time-discretization of geodesics

In a sequence of papers, Rumpf and Wirth [Wir09, WBRS09, RW13, RW15] have introduced a time-discrete analogon of the continuous geodesic calculus presented in the previous section. The resulting time-discrete geodesic calculus has already been applied to several Riemannian manifolds or shape spaces, *e.g.* in [HRWW12, HRS$^+$14, BER15, MRSS15, Per15]. In this section, we provide a survey of the time-discrete geodesic calculus proposed by Rumpf and Wirth and present () important convergence results. As before, we start with a variational formulation of *discrete* geodesics, defined via minimizers of a *discrete* path energy[5].
In the continuous setting the starting point of a geometric calculus on a Riemannian manifold is usually the definition of a Riemannian metric. However, as we will see, the discrete geodesic calculus is solely based on the notion of a (squared) Riemannian distance resp. a local approximation thereof. Obviously, a Riemannian distance is induced by the metric (*cf*. eq. (3.3)). On the other hand, given the Riemannian distance dist, one can recover the Riemannian metric g_p at some point $p \in \mathcal{M}$ by

$$g_p(V,W) = \frac{1}{2}\,\partial_2^2\,\mathrm{dist}^2(p,p)(V,W)\,, \quad V,W \in T_p\mathcal{M}\,. \tag{3.7}$$

For many applications, *e.g.* when dealing with physical shape spaces, it is difficult to define a Riemannian metric a priori. On the other hand, it is often much easier to come up with the notion of a distance, *e.g.* by using a physically sound dissimilarity measure (*cf*. Sec. 2). To account for this circumstance as well as for the fact that Riemannian distances are in practice hard to compute (as they require solving an optimization problem), the discrete geodesic calculus is actually based on an *approximation* of the squared Riemannian distance which is easy to evaluate and consistent with the metric by definition due to (3.7).

Variational time-discretization. In the following, we denote an ordered set of points $Y^K = (y_0, \dots, y_K)$ in the manifold $\mathcal{M}$ as a time-discrete K-path. Often we interpret this discrete path as a uniform sampling of a smooth curve $y : [0,1] \to \mathcal{M}$, *i.e.* we have $y_k = y(t_k)$ with $t_k = k\tau$ for $k = 0, \dots, K$ where $\tau = K^{-1}$ and $K \in \mathbb{N}$ denotes the sample size. Instead of using a straightforward time-discretization of the continuous path energy (3.2) we first consider the following estimates

$$\mathcal{L}[(y(t))_{t\in[0,1]}] \geq \sum_{k=1}^{K} \mathrm{dist}(y_{k-1}, y_k)\,, \qquad \mathcal{E}[(y(t))_{t\in[0,1]}] \geq \frac{1}{\tau}\sum_{k=1}^{K} \mathrm{dist}^2(y_{k-1}, y_k)\,, \tag{3.8}$$

where equality holds for geodesic paths due to the constant speed property. The first estimate is straightforward, and the second estimate follows with the Cauchy-Schwarz inequality, *i.e.*

$$\sum_{k=1}^{K} \mathrm{dist}^2(y_{k-1}, y_k) \leq \sum_{k=1}^{K} \left(\int_{(k-1)\tau}^{k\tau} \sqrt{g_{y(t)}(\dot y(t), \dot y(t))}\ \mathrm{d}t \right)^2 \leq \sum_{k=1}^{K} \tau \int_{(k-1)\tau}^{k\tau} g_{y(t)}(\dot y(t), \dot y(t))\ \mathrm{d}t\,,$$

since the expression on the right hand side is exactly $\tau\ \mathcal{E}[(y(t))_{t\in[0,1]}]$.

The estimate on the path energy in (3.8) suggest that the sum on the right hand side might be a reasonable approximation of $\mathcal{E}$. However, as already mentioned in the beginning, the squared Riemannian distance dist^2 is often difficult to compute in practice. Therefore we assume there is a functional $\mathcal{W} : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ which is a *local* approximation of the squared Riemannian distance dist^2. In detail, it is supposed that $\mathcal{W}$ is weakly lower semi-continuous in both arguments and that it satisfies the following hypotheses:

(H2)
$$\begin{cases} \text{There exist } \varepsilon, C > 0 \text{ such that for all } y,\ \tilde y \in \mathcal{M}\text{:} \\ \mathrm{dist}(y,\tilde y) \leq \varepsilon \ \Rightarrow\ |\mathcal{W}[y,\tilde y] - \mathrm{dist}^2(y,\tilde y)| \leq C\mathrm{dist}^3(y,\tilde y) \\ \mathcal{W} \text{ is coercive in the sense } \mathcal{W}[y,\tilde y] \geq y(\mathrm{dist}(y,\tilde y)) \\ \text{for a strictly increasing, continuous function } y \text{ with } y(0) = 0 \text{ and } \lim_{d\to\infty} y(d) = \infty. \end{cases}$$

The first property says there is a $C > 0$, such that $|\,\mathcal{W}[y,\tilde y] - \mathrm{dist}^2(y,\tilde y)\,| \leq C\mathrm{dist}^3(y,\tilde y)$. Note that $\mathcal{W}$ is not required to be symmetric. For g smooth enough, a valid approximation of dist^2 is *e.g.* given by $\mathcal{W}[y,\tilde y] = \frac{1}{2}g_y(\tilde y - y, \tilde y - y)$. In general, the following theorem states that $g_y = \frac{1}{2}\mathcal{W}_{,22}[y,y]$ implies (H2) for smooth g and $\mathcal{W}$:

[5] We will often omit the prefix "time" when referring to a time-discrete object.

Theorem 3.8 (Consistency conditions, [RW15])**.** *Unter hypothesis* (H1)*, if $\mathcal{W}$ is twice Gâteaux - differentiable on $\mathcal{M}\times\mathcal{M}$ with bounded second Gâteaux derivative, then $\mathcal{W}[y,\tilde{y}] = \mathrm{dist}^2(y,\tilde{y}) + O(\mathrm{dist}^3(y,\tilde{y}))$ for $\tilde{y}$ close to $y \in \mathring{\mathcal{M}}$ implies*

$$\mathcal{W}[y,y] = 0\,, \quad \mathcal{W}_{,2}[y,y](V) = 0\,, \quad \mathcal{W}_{,22}[y,y](V,W) = 2g_y(V,W)$$

for any $V, W \in \mathbf{V}$. Furthermore, $\mathcal{W}_{,1}[y,y](V) = 0$ and

$$\mathcal{W}_{,11}[y,y](V,W) = -\mathcal{W}_{,12}[y,y](V,W) = -\mathcal{W}_{,21}[y,y](V,W) = \mathcal{W}_{,22}[y,y](V,W)\,.$$

If $\mathcal{W}$ is even three times Fréchet-differentiable, the implication becomes an equivalence.

*The proof of Thm. 3.8 can also be found in [RW15, Lemma 4.6].

We arrive at the following definition of a *discrete path energy* and a *discrete path length* (see [RW15]):

Definition 3.9 (Discrete length and energy)**.** For a discrete K-path $Y^K = (y_0, \dots, y_K)$ with $y_k \in \mathcal{M}$ for $k = 0, \dots, K$ we define the *discrete length* L^K and the *discrete energy* E^K by

$$L^K[Y^K] = \sum_{k=1}^{K} \sqrt{\mathcal{W}[y_{k-1}, y_k]}\,, \qquad E^K[Y^K] = K \sum_{k=1}^{K} \mathcal{W}[y_{k-1}, y_k]\,. \tag{3.9}$$

Then a *discrete geodesic* (of order K) is defined as a minimizer of $E^K[Y^K]$ for fixed end points y_0, y_K.

Existence and uniqueness of discrete geodesics. Next we prove the existence and uniqueness of discrete geodesics as presented in [RW15]. The proofs follow the same ideas as the corresponding continuous theorems presented in the previous subsection.

Theorem 3.10 (Existence of discrete geodesics, [RW15])**.** *Given y_A, $y_B \in \mathcal{M}$, there is a discrete geodesic path $(y_0, \dots, y_K)$ which minimizes the discrete energy E^K over all discrete paths $(\tilde{y}_0, \dots, \tilde{y}_K)$ with $\tilde{y}_0 = y_A$ and $\tilde{y}_K = y_B$.*

Proof: see Thm 4.3 in [RW15]. □

Theorem 3.11 (Uniqueness of discrete geodesics, [RW15])**.** *Let* (H1) *and* (H2) *hold and assume $\mathcal{W}$ to be twice Fréchet-differentiable on $\mathcal{M} \times \mathcal{M}$. For all $y_A \in \mathring{\mathcal{M}}$ and $K \in \mathbb{N}$ there exists $\varepsilon > 0$ such that there exists a unique discrete geodesic $(y_0, \dots, y_K)$ with $y_0 = y_A$ and $y_K = y_B$ for all y_B with $\|y_A - y_B\|_{\mathbf{V}} < \varepsilon$.*

Proof: see Thm 4.7 in [RW15]. □

***Properties of discrete path energy and discrete geodesics.** We state further properties from [RW15]:

Theorem [Convergence of path energy, [RW15]]. *Under hypothesis* (H2) *there exists $\delta > 0$ such that* $\mathrm{dist}(y_A, y_B) < \sqrt{K}\delta$ *implies*

$$\Big| \min_{\substack{(y_0,\dots,y_K)\\ y_0 = y_A, y_K = y_B}} E^K[(y_0, \dots, y_K)] - \mathrm{dist}^2(y_A, y_B) \Big| =\leq \frac{C'}{K} \mathrm{dist}^3(y_A, y_B)\,.$$

Theorem [Equidistribution of points along discrete geodesics, [RW15]]. *Under hypothesis* (H2) *there exists $\delta > 0$ such that if* $\mathrm{dist}(y_A, y_B) < \sqrt{K}\delta$*, then discrete geodesics satisfy* $\mathrm{dist}(y_{k-1}, y_k) \leq C\tau$ *for all $k = 1, \dots, K$ with the constant $C > 0$ only depending on* $\mathrm{dist}(y_A, y_B)$.

Convergence. In this paragraph we prove convergence of the discrete path energy to the continuous path energy in the sense of Γ-convergence (as defined in Def.2.1). As a direct consequence this implies the convergence of discrete geodesics to continuous ones. However, the discrete path energy E^K is defined on a finite sequence of points in $(\mathcal{M})^K$, whereas the continuous path energy $\mathcal{E}$ is defined on continuous paths $(y(t))_{t\in[0,1]}$ on $\mathcal{M}$. To this end, we extend the definition of the discrete path energy to continuous paths on $\mathcal{M}$ which are composed of shortest geodesic segments.

For the remainder of this subsection we set $t_k = k/K$ and $y_k = y(t_k)$ for a given path $y : [0,1] \to \mathcal{M}$ and $k = 0, \dots, K$. We define

$$\mathcal{A}^K := \left\{ y \in L^2\left((0,1), \mathcal{M}\right) \,:\, y\Big|_{[t_{k-1},t_k]} \text{ is shortest geodesic segment} \right\}.$$

That means, if $y \in \mathcal{A}^K$, then y is piecewise geodesic interpolation of the points $y_0, \dots, y_K$. Now we define an energy $\hat{E}^K : L^2((0,1); \mathbf{Y}) \to \mathbb{R}$ via

$$\hat{E}^K[y] := \begin{cases} E^K[(y_0, \dots, y_K)] & , \text{ if } y \in \mathcal{A}^K \\ \infty & , \text{ else} \end{cases}$$

Theorem 3.12 (Γ-convergence of the discrete energy, [RW15])**.** *Assuming* (H1) *and* (H2)*, $\hat{E}^K$ converges to $\mathcal{E}$ for $K \to \infty$ in the sense of Γ-convergence wrt. the $L^2((0,1); \mathbf{Y})$-topology.*

Proof: See proof of [RW15, Thm. 4.8]. □

Now let $(y_0, \dots, y_K)$ be a discrete geodesic, i.e. a minimizer of E^K. Let y^K be the piecewise geodesic interpolation of $y_0, \dots, y_K$. From the proof of the previous theorem we know $\mathcal{E}[y^K] \le \bar{E} < \infty$. The the proof of Thm. 3.2 implies that there is a subsequence of $(y^K)_K$, still denoted by $(y^K)_K$, that with $y^K \to y$ in $C^0([0,1], \mathbf{Y})$. Due to the properties of Γ-convergence discussed in Sec. 2.2 we get (compare to [RW15, Corollary 4.9]):

Corollary 3.13 (Convergence of discrete geodesics)**.** *Minimizers of the discrete path energy E^K, which are piecewise geodesically interpolated, converge to minimizers of the continuous path energy $\mathcal{E}$ in $C^0([0,1], \mathbf{Y})$.*

Remark (i): Taking into account the equivalence of the $\mathbf{V}$ topology and the manifold topology, a similar argument can be given for the piecewise linear interpolation y^K_{lin} of discrete geodesics instead of piecewise geodesic interpolations so that the above convergence obviously also holds for y^K_{lin}.

Remark (ii): Rumpf and Wirth [RW15] obtain even stronger convergence estimates under additional smoothness hypotheses (in addition to (H1) and (H2)). In detail, they prove that the convergence in $C^0([0,1]; \mathbf{Y})$ ensured by Thm. 3.12 is actually much stronger with velocities converging in $L^2((0,1); \mathbf{V})$.

Remark (iii): Note that discrete minimizers of the discrete path length L^K are in general unrelated to continuous geodesics. Let us consider the case $\mathcal{M} = \mathbb{R}^2 \setminus B_r$, where $B_r = \{x : |x| < r\}$, and $\mathcal{W}[y, \tilde{y}] = \|y - \tilde{y}\|^2$, as depicted in Fig. 5. Then a discrete path $(y_0, \dots, y_K)$ connecting $y_A = (-\alpha r, 0)$ and $y_B = (\alpha r, 0)$, $\alpha > 1$, that minimizes the time-discrete path energy tend to distribute uniformly along the connecting curve. If $r \gg \operatorname{dist}(y_A, y_B)/K$ this is not realizable along a straight line connecting y_A and y_B, *cf*. Fig. 5. However, the distribution of points representing a discrete minimizer of L^K is arbitrary, since moving points along the connecting line does not alter the length.

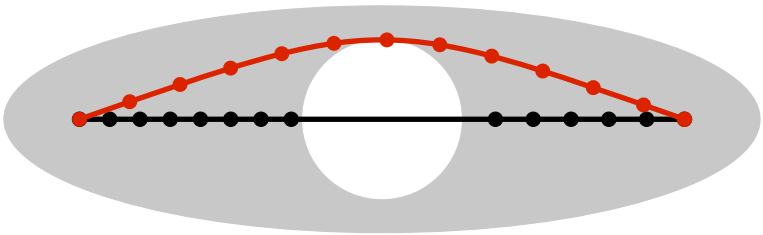


Figure 5: In general, minimizers of the discrete path length do not converge to continuous geodesics.

3.3 Riemannian splines

For two points $p_1, p_2 \in \mathcal{M}$ a smooth interpolation $y : [0,1] \to \mathcal{M}$ with $y(0) = p_1$ and $y(1) = p_2$ is given by the connecting geodesic path. However, for a sequence $0 = t_1 < t_2 < \ldots < t_J = 1$ and corresponding points $p_1, \ldots, p_J \in \mathcal{M}$, there is in general no geodesic curve $y : [0,1] \to \mathcal{M}$ that fulfills the *interpolation constraints*

$$y(t_j) = p_j\,, \quad j = 1, \ldots, J\,. \tag{3.10}$$

In particular, a curve y satisfying (3.10) does in general not comply with the geodesic equation $\frac{D}{dt}\dot{y} = 0$. For example, a piecewise geodesic curve connecting $p_1, \ldots, p_J$ fulfills $\frac{D}{dt}\dot{y} = 0$ on each segment (t_j, t_{j+1}), $j = 1, \ldots, J-1$, but exhibits discontinuities in $\dot{y}$ at the interpolation points. Nevertheless, if one is interested in a curve that on the one hand satisfies the interpolation constraints *exactly* and on the other hand is *as smooth as possible*, one might consider the geodesic equation as a *penalty term*. This motivation leads to the functional

$$\mathcal{F}[(y(t))_{t\in[0,1]}] = \int_0^1 g_{y(t)}\left(\frac{D}{dt}\dot{y}(t), \frac{D}{dt}\dot{y}(t)\right)\,\mathrm{d}t\,, \tag{3.11}$$

where $\frac{D}{dt}$ denotes the covariant derivative along y as defined by Def. 3.6. The generaliuzed multiple interpolation task is then given by

$$\text{Compute} \quad y \in \operatorname*{arg\,min}_{y:[0,1]\to\mathcal{M}} \mathcal{F}[y] \quad \text{subject to (3.10)}\,.$$

In the finite dimensional Euclidean setting, *i.e.* $\mathcal{M} = \mathbb{R}^d$ and g_p denotes the standard Euclidean product, the covariant derivative of $\dot{y}$ is simply given by the second time derivative $\ddot{y}$, *i.e.* we have

$$\mathcal{F}_{\mathrm{Euc}}[(y(t))_{t\in[0,1]}] = \int_0^1 \|\ddot{y}(t)\|^2\,\mathrm{d}t\,. \tag{3.12}$$

Consider a discretization of the unit interval $I = [0,1]$ with nodes $I_h = \{0 = z_0 < z_2 < \ldots < z_N = 1\}$. A spline function of degree k on I_h is a function $s \in C^{k-1}(I, \mathbb{R}^d)$ such that s is a polynomial of degree $\leq k$ on each interval $[z_{n-1}, z_n]$, $n = 1, \ldots, N$. The following theorem is often refered to as Schoenberg's theorem although it has been proved first by de Boor[6]:

Theorem [de Boor, 1963]. *For $0 = t_1 < t_2 < \ldots < t_J = 1$ and $p_1, \ldots, p_J \in \mathbb{R}^d$ there is a unique minimizer $y \in C^2([0,1], \mathbb{R}^d)$ of $\mathcal{F}_{\mathrm{Euc}}$ that satisfies the interpolation constraints $y(t_j) = p_j$ for $j = 1, \ldots, J$ as well as one of the boundary conditions*

$$\begin{aligned}
&\ddot{y}(0) = \ddot{y}(1) = 0\,, && \text{(natural b.c.)}\\
&\dot{y}(0) = v_0, \quad \dot{y}(1) = v_1 \quad \text{for given } v_0, v_1 \in \mathbb{R}^d\,, \text{ or} && \text{(Hermite b.c.)}\\
&y(0) = y(1), \quad \dot{y}(0) = \dot{y}(1), \quad \ddot{y}(0) = \ddot{y}(1)\,. && \text{(periodic b.c.)}
\end{aligned}$$

The minimizer is given by the unique cubic spline, i.e. *a spline of degree 3, satisfying the interpolation constraints and boundary conditions.*

As (3.11) can be seen as a generalization of (3.12) to Riemannian manifolds, we refer to $\mathcal{F}$ as *spline energy* and we denote minimizers of $\mathcal{F}$ as *Riemannian (cubic) splines*. The boundary condition then read as follows:

$$\begin{aligned}
&\frac{D}{dt}\dot{y}(0) = \frac{D}{dt}\dot{y}(1) = 0\,, && \text{(natural b.c.)}\\
&\dot{y}(0) = v_0, \quad \dot{y}(1) = v_1 \quad \text{for given } v_0 \in T_{y(0)}\mathcal{M}\,,\ v_1 \in T_{y(1)}\mathcal{M}\,, \text{ or} && \text{(Hermite b.c.)}\\
&y(0) = y(1), \quad \dot{y}(0) = \dot{y}(1), \quad \frac{D}{dt}\dot{y}(0) = \frac{D}{dt}\dot{y}(1)\,. && \text{(periodic b.c.)}
\end{aligned} \tag{3.13}$$

Although the above theorem states the existence of interpolating splines in the Euclidean space, this is not true for general manifolds (cf. [HRW17, Lemma 2.15]):

[6] Actually, Schoenberg cites de Boor's paper [dB63] when referring to this result in [Sch64b]. F or further reading on this we refer to [Sch73, Sch64a], a simple proof is given *e.g.* in [DH02, 7.4].

Lemma 3.14 (Nonexistence of Riemannian splines, [HRW17])**.** *Let $\mathcal{M}$ be any manifold with a closed geodesic curve C and a point $\bar{y}_1 \in C$ such that any locally geodesic curve connecting $\bar{y}_1$ with itself lies inside C. Then, minimizers of $\mathcal{F}$ under the interpolation contraints* (3.10) *do not exist in general.*

Proof: It suffices to provide a counterexample. Consider $\mathcal{M}$ to be a cylinder in $\mathbb{R}^3$ of infinite length and perimeter 1. Let $t_1 = 0$, $t_2 = r \in (0,1) \setminus \mathbb{Q}$, $t_3 = 1$, and choose an arbitrary point $p_1 = p_3 \in \mathcal{M}$. Let $p_2 \in \mathcal{M}$ be the point opposite p_1. Now define $\xi : \mathbb{R} \to \mathcal{M}$ to be a 1-periodic arclength-parameterization of the circle through p_1, p_2, p_3 with $\xi(0) = p_1$. In particular we have $\xi(\mathbb{N}) = p_1$ and $\xi(\mathbb{N} + 1/2) = p_2$. For arbitrary $m, n \in \mathbb{N}$ consider the Euclidean cubic spline $x : [0,1] \to \mathbb{R}$ with $x(t_1) = x_1 = 0$, $x(t_2) = x_2 = m + \frac{1}{2}$, and $x(t_3) = x_3 = n$. Obviously, $y = \xi \circ x$ is a curve on $\mathcal{M}$ satisfying (3.10). Furthermore, its spline energy can be computed explicitly as

$$\mathcal{F}[y] = \int_0^1 |\ddot{y}(t)|^2 \,\mathrm{d}t = \int_0^1 |\ddot{x}(t)|^2 \,\mathrm{d}t = \frac{3((x_2 - x_1)(t_3 - t_1) - (x_3 - x_1)(t_2 - t_1))^2}{(t_3 - t_2)^2 (t_3 - t_1)(t_2 - t_1)^2} = \frac{3(m + \frac{1}{2} - rn)^2}{(1-r)^2 r^2},$$

where we used $\ddot{y}(t) = \frac{d}{dt}(\dot{\xi}(x(t)) \cdot \dot{x}(t)) = 0 \cdot \dot{x}(t) + 1 \cdot \ddot{x}(t)$. By Dirichlet's approximation theorem there exist $m, n \in \mathbf{Z}$ that make the above arbitrarily small, hence we have $\inf \mathcal{F}[y] = 0$.
However, there is no curve y with $\mathcal{F}[y] = 0$. Indeed, such a curve would satisfy $\frac{D}{\mathrm{d}t}\dot{y} = 0$, which on the cylinder results in a regular helix with constant speed. Since $y(0) = y(1)$, the helix is degenerate and winds round the circle at constant speed so that necessarily $y(t) = \xi(mt)$ for some $m \in \mathbf{Z}$. However, the preimage of p_2 under y does not contain r so that (3.10) is violated, $r \notin y^{-1}(p_2) = \{\frac{1}{2m}, \frac{3}{2m}, \ldots, \frac{2m-1}{2m}\}$. □

Remark: This construction can easily be transferred onto a general manifold with a closed geodesic, where the circle is replaced by the closed geodesic and the interpolation conditions are chosen correspondingly. Indeed, the infimum of the spline energy on an analogous sequence of curves, now mapping onto the closed geodesic, vanishes. Hence, any minimizer, if it exists, must be a (local) geodesic characterized by $\frac{D}{\mathrm{d}t}\dot{y} = 0$ and fulfilling the interpolation conditions. The above argument shows that this is impossible.

The counterexample above relies on the fact that $\mathcal{F}$ only penalizes *variations* in the velocity and not the velocity itself. In particular, it is zero whenever the velocity is constant and it is not sensitive to the total length of the curve. To account for this circumstence we will instead consider $\mathcal{F} + \sigma\mathcal{E}$ for some $\sigma > 0$. Now the total length of the curve is penalized as well and a construction as in Lemma 3.14 can no longer be energetically advantageous. However, in order to obtain existence we have to specify an admissible Riemannian manifold first. Let $\mathbf{V}$ be a separable Hilbertspace (which is reflexive), $\mathbf{Y}$ a Banach space, such that $\mathbf{V}$ compactly embeds into $\mathbf{Y}$, and set $\mathcal{M} := \mathbf{V}$. Next, we have to define an admissible metric:

Definition 3.15 (Admissible metric)**.** A Riemannian metric $g : \bigcup_{y \in \mathcal{M}} (\{y\} \times T_y\mathcal{M} \times T_y\mathcal{M}) \to \mathbb{R}$ on $\mathcal{M}$ shall be called *admissible* if it can be extended to a function $g : \mathbf{Y} \times \mathbf{V} \times \mathbf{V} \to \mathbb{R}$ of the form

$$g_y(v, w) = g^c_y(v, w) + \mathcal{Q}(v, w) \tag{3.14}$$

for some compact part g^c, which depends on the position y, and a quadratic part $\mathcal{Q}$, where the following hypotheses shall be satisfied for all $v \in \mathbf{V}$.

(i) g^c is symmetric in the last two arguments and $g^c_y(v, v) \leq C^* \|v\|^2_{\mathbf{Y}}$ for some constant C^*.

(ii) g^c is twice differentiable with bounded derivatives as a function $g^c : \mathbf{Y} \to \mathbf{Y}' \otimes \mathbf{Y}'$.

(iii) $\mathcal{Q}$ is symmetric positive semi-definite and bilinear on $\mathbf{V} \times \mathbf{V}$ with $\mathcal{Q}(v, v) \leq C^{**} \|v\|^2_{\mathbf{V}}$ for some constant C^{**}.

(iv) g is uniformly coercive with respect to the $\mathbf{V}$ norm, *i.e.* $c^* \|v\|^2_{\mathbf{V}} \leq g_y(v, v)$ for some $c^* > 0$.

Here, $\mathbf{V}'$ and $\mathbf{Y}'$ denote the dual spaces to $\mathbf{V}$ and $\mathbf{Y}$, and $\mathbf{V}' \otimes \mathbf{V}'$ and $\mathbf{Y}' \otimes \mathbf{Y}'$ are equipped with the topology induced by the injective cross norm.

Remark I: As a direct consequence of (ii) there exists a strictly increasing continuous function β with $\beta(0) = 0$ such that

$$|g^c_y(v, v) - g^c_{\tilde{y}}(v, v)| \leq \beta(\|y - \tilde{y}\|_{\mathbf{Y}}) \|v\|^2_{\mathbf{Y}} \text{ for all } y,\ \tilde{y} \in \mathbf{Y} \text{ and all } v \in \mathbf{V}.$$

Remark II: Property (i) is stronger than the corresponding assumption in (H1), i.e. $g_y(v,v) \leq C^* \|v\|_{\mathbf{V}}^2$, since $\|.\|_{\mathbf{Y}} \leq C \|.\|_{\mathbf{V}}$ due to the bounded embedding. This $\|.\|_{\mathbf{V}}$-boundedness is, however, admissible for the quadratic part $\mathcal{Q}$, which does not depend on the position $y \in \mathcal{M}$ on the other hand.

Remark III: The decomposition of g into g^c and $\mathcal{Q}$ will be necessary to establish weak continuity of the associated Christoffel operator (see Lemma 3.16), which is required for an existence result of Riemannian splines. For a general g_y, which is uniformly bounded and positive definite on $\mathbf{V} \times \mathbf{V}$ for all $y \in \mathbf{Y}$ but nonlinear in y, the right-hand side of

$$2g_y(\Gamma_y(v,w),z) = (D_y g_y)(w)(v,z) - (D_y g_y)(z)(v,w) + (D_y g_y)(v)(w,z) \tag{3.15}$$

will in general not be weakly continuous jointly in v and w on infinite-dimensional spaces $\mathbf{V}$.

Before, we are able to prove existence of minimizers of $\mathcal{F} + \sigma\mathcal{E}$ with $\sigma > 0$, we have to establish a weak continuity result for the Christoffel operator, which will imply weak lower semi-continuity of the (augmented) spline energy.

Lemma 3.16 (Weak continuity of the Christoffel operator, [HRW17])**.** *On an admissible Riemannian manifold $(\mathcal{M}, g)$ the Christoffel operator is weakly continuous in the sense*

$$\Gamma_{y_k}(v_k, w_k) \to \Gamma_y(v,w) \text{ in } \mathbf{V} \text{ as } k \to \infty$$

for $y_k \to y$ strongly in $\mathbf{Y}$ and $(v_k, w_k) \rightharpoonup (v,w)$ weakly in $\mathbf{V} \times \mathbf{V}$. In more detail,

$$\begin{aligned} &\|\Gamma_{y_k}(v_k,w_k) - \Gamma_y(v,w)\|_{\mathbf{V}} \\ &\leq C\left(\|w_k\|_{\mathbf{Y}}\|v_k - v\|_{\mathbf{Y}} + \|w_k - w\|_{\mathbf{Y}}\|v\|_{\mathbf{Y}} + \|y_k - y\|_{\mathbf{Y}}\|w\|_{\mathbf{V}}\|v\|_{\mathbf{V}} + \|y_k - y\|_{\mathbf{Y}}\|\Gamma_{y_k}(v_k,w_k)\|_{\mathbf{Y}}\right) \end{aligned} \tag{3.16}$$

for some constant $C > 0$ only depending on c^* and the derivative bounds on g^c from Definition 3.15(ii).

Proof: See [HRW17, Lemma 2.17]. □

Next, we prove the lower semi-continuity property of $\mathcal{F}$ in the weak H^2-topology:

Lemma 3.17 (Continuity properties of spline energy, [HRW17])**.** *For $(\mathcal{M}, g)$ admissible, the spline energy $\mathcal{F}$ is lower semi-continuous wrt. the weak $H^2((0,1);\mathbf{V})$-topology.*

Proof: See [HRW17, Lemma 2.18]. □

Now we can finally prove the existence of interpolating splines as minimizers of the augmented/regularized functional $\mathcal{F} + \sigma\mathcal{E}$:

Theorem 3.18 (Existence of spline interpolations, [HRW17])**.** *For $\sigma > 0$ and $(\mathcal{M}, g)$ admissible there exists a minimizer y of $\mathcal{F}^\sigma[y] := \mathcal{F}[y] + \sigma\mathcal{E}[y]$ subject to* (3.10) *under natural, Hermite, or periodic boundary conditions in the Sobolev space $H^2((0,1);\mathbf{V})$.*

Sketch of proof: 1) First, we show that the regularized spline energy $\mathcal{F}^\sigma$ with condition (3.10) is coercive in $H^2 = H^2((0,1);\mathbf{V})$. Indeed, let $\mathcal{F}[y] + \sigma\mathcal{E}[y] < M$ for some $M \in \mathbb{R}$, then $\|\dot{y}\|_{L^2}^2 \leq \frac{\mathcal{E}[y]}{c^*} \leq \frac{M}{\sigma c^*}$, and by (3.10) and Poincaré's inequality it follows that $\|y\|_{H^1}^2 \leq C(M)$. Furthermore, using the reverse triangle inequality and Young's inequality we have (abbreviating $\Gamma = \Gamma_y(\dot{y},\dot{y})$)

$$\begin{aligned} \tfrac{M}{c^*} \geq \tfrac{1}{c^*}\mathcal{F}[y] \geq \int_0^1 \|\frac{D}{\mathrm{d}t}\dot{y}\|_{\mathbf{V}}^2 \,\mathrm{d}t &\geq \int_0^1 \|\ddot{y}\|_{\mathbf{V}}^2 - 2\|\ddot{y}\|_{\mathbf{V}}\|\Gamma\|_{\mathbf{V}} + \|\Gamma\|_{\mathbf{V}}^2 \,\mathrm{d}t \\ &\geq \int_0^1 \tfrac{1}{2}\|\ddot{y}\|_{\mathbf{V}}^2 - \|\Gamma\|_{\mathbf{V}}^2 \,\mathrm{d}t \geq \tfrac{1}{2}\|\ddot{y}\|_{L^2}^2 - \left(\tfrac{3}{2}\tfrac{\|D_y g^c\|}{c^*}\right)^2 \|\dot{y}\|_{L^4}^4, \end{aligned}$$

where in the last inequality we used the estimate $\|\Gamma_y(\dot{y},\dot{y})\|_{\mathbf{V}} \leq \frac{3}{2}\frac{\|D_y g^c\|}{c^*}\|\dot{y}\|_{\mathbf{Y}}^2$ (cf. (3.15)). From this one can deduce that $\|\ddot{y}\|_{L^2}$ is bounded by a constant depending solely on M, which is proved by contradiction (by means of a decreasing rearrangement, for details we refer to [HRW17, Thm. 2.19]).

2) As a consequence of the coercivity, a minimizing sequence $(y_k)_{k=1,\dots}$ is uniformly bounded in H^2, and by the reflexivity of the space H^2 we obtain a weakly converging subsequence, again denoted by $(y_k)_{k=1,\dots}$, which converges to some y in H^2. Finally, the weak lower semi-continuity of $\mathcal{F}$ by Lemma 3.17 and $\mathcal{E}$ by Thm. 3.2 implies

$$\mathcal{F}^\sigma[y] \leq \liminf_{k\to\infty} \mathcal{F}^\sigma[y_k]\,.$$

Furthermore, the weak limit y satisfies (3.10) and the boundary conditions since they are continuous with respect to weak convergence in $H^2((0,1);\mathbf{V})$. Thus, y is the sought spline interpolation. □

Remark: One can also prove that y is in $C^{2,\frac{1}{2}}([\delta,1-\delta];\mathbf{V})$ for every $\delta \in (0,\frac{1}{2})$, cf. [HRW17, Thm. 2.19].

3.4 Variational time-discretization of Riemannian splines

In this section, we derive a consistent time-discretization of the spline energy (3.11) that fits into the framework of time-discrete geodesic calculus presented in Sec. 3.2.

As a motivation we start taking a look at the Euclidean setup, *i.e.* $\mathcal{M} = \mathbb{R}^d$ with the standard Euclidean scalar product $g_p =< .,. >$ and $\mathcal{W}[y,\tilde{y}] = \|y-\tilde{y}\|^2$. We consider a curve $y : [0,1] \to \mathcal{M}$ and for some stepsize $\tau = K^{-1}$ a uniform sampling $y_k = y(t_k)$ with $t_k = k\tau$ for $k = 0,\dots,K$. Then we have

$$\mathcal{E}_{Euc}[y] = \int_0^1 \|\dot{y}(t)\|^2\,\mathrm{d}t \approx \sum_{k=1}^K \left\|\dot{y}\left(\frac{t_{k-1}+t_k}{2}\right)\right\|^2 \approx \sum_{k=1}^K \tau \left\|\frac{y_{k-1}-y_k}{\tau}\right\|^2 = K\sum_{k=1}^K \mathcal{W}[y_{k-1},y_k]\,.$$

Now the covariant derivative of $\dot{y}$ is simply the second time derivative $\ddot{y}$. Approximating the integrand of the Euclidean spline energy $\int \|\ddot{y}\|^2\,\mathrm{d}t$ by a second order finite difference quotient yields

$$\|\ddot{y}(t_k)\|^2 \approx \left\|\frac{2\,y_k - y_{k-1} - y_{k+1}}{\tau^2}\right\|^2 = 4\tau^{-4}\left\|y_k - \frac{y_{k-1}+y_{k+1}}{2}\right\|^2.$$

Using this approximation as well as natural bounday conditions $\ddot{y}(0) = \ddot{y}(1) = 0$ we arrive at

$$\mathcal{F}_{Euc}[y] = \int_0^1 \|\ddot{y}(t)\|^2\,\mathrm{d}t \approx \sum_{k=1}^{K-1} \tau\,\|\ddot{y}(t_k)\|^2 \approx 4\tau^{-3}\sum_{k=1}^{K-1}\left\|y_k - \frac{y_{k-1}+y_{k+1}}{2}\right\|^2 = 4K^3\sum_{k=1}^K \mathcal{W}[y_k,\tilde{y}_k]\,,$$

with $\tilde{y}_k = \frac{1}{2}(y_{k-1}+y_{k+1})$. The key insight is to interprete the local average $\tilde{y}_k$ as the midpoint of a geodesic in the Euclidean space connecting y_{k-1} and y_{k+1}, where a geodesic is given by the straight connecting line, i.e.

$$\tilde{y}_k = \operatorname*{arg\,min}_y \left(\mathcal{W}[y_{k-1},y] + \mathcal{W}[y,y_{k+1}]\right)\,.$$

Replacing the squared Euclidean distance by a general approximation functional $\mathcal{W}$, and the local average $\tilde{y}_k$ by the midpoint of a short geodesic connecting y_{k-1} and y_{k+1}, one obtains

Definition 3.19 (Discrete spline energy)**.** For $K \in \mathbb{N}$ let $Y^K = (y_0,\dots,y_K)$ be a discrete K-path in $\mathcal{M}$. We define the discrete spline energy by

$$F^K[Y^K] = 4K^3\sum_{k=1}^{K-1}\mathcal{W}[y_k,\tilde{y}_k]\,, \tag{3.17}$$

subject to the constraint that $(y_{k-1},\tilde{y}_k,y_{k+1})$ is a discrete geodesic for $k = 1,\dots,K-1$, *i.e.*

$$\tilde{y}_k = \operatorname*{arg\,min}_{y\in\mathcal{M}}\Big(\mathcal{W}[y_{k-1},y] + \mathcal{W}[y,y_{k+1}]\Big)\,,\ \text{for } k = 1,\dots,K-1\,. \tag{3.18}$$

Note that the natural boundary conditions have been incorporated implicitly since the summation in (3.17) goes from $k=1$ to $k=K-1$ assuming that there is no contribution at the boundary. However, corresponding Hermite or periodic boundary conditions can be chosen as well. Furthermore, we can also pose discrete interpolation constraints. Let $0 = k_1 < k_2 < \dots < k_I = K$ be an index set ($I \geq 2$) and let $p_1,\dots,p_I \in \mathcal{M}$ be given. We say that Y^K is a discrete spline interpolating $y_{k_1},\dots,y_{k_I}$ if it minimizes (3.17) subject to

$$y_{k_i} = p_i\,,\quad i = 1,\dots,I\,. \tag{3.19}$$

If $I = 2$, *i.e.* we only fix y_0 and y_K, the discrete spline is precisely the discrete geodesic connecting y_0 and y_K.

Remark: As in the contiuous case, existing of minimizers of F^K subject to (3.19) and appropiate boundary conditions, can only be shown for the augmented/regularized discrete functional

$$F^K[y] + \sigma\, E^K[y]\,, \quad \sigma > 0\,.$$

Additionally we need a splitting of the approximative functional $\mathcal{W}$ that represents the splitting of the metric as defined in Def. 3.15, i.e. $\mathcal{W}[y,\tilde{y}] = \mathcal{W}^c[y,\tilde{y}] + \mathcal{Q}(\tilde{y}-y, \tilde{y}-y)$. Indeed, without this assumption on the functional $\mathcal{W}$ one cannot expect the discrete regularized spline energy to possess minimizers in general. In fact, if one considers a minimizing sequence $\big((y_0^j,\dots,y_K^j)\big)_{j=1,\dots}$ in $\mathcal{M}^{K+1}$, the coercivity of $\mathcal{W}$ only leads to weak convergence in $\mathbf{V}^{K+1}$ for a subsequence. However, weak convergence of y_{k-1}^j and y_{k+1}^j as $j\to\infty$ does not necessarily imply weak convergence of their geodesic midpoint for general functionals $\mathcal{W}$ obeying only the hypothesis posed in [RW15]. Thus, $\mathcal{W}[y_k^j, \widetilde{y}_k^j]$ may not be lower semi-continous as $j\to\infty$, preventing the existence of a minimizer. Indeed, $4K^4\mathcal{W}[y_k^j, \widetilde{y}_k^j]$ is the discrete counterpart of $g_y(\frac{D}{\mathrm{d}t}\dot{y}, \frac{D}{\mathrm{d}t}\dot{y})$, and thus the lack of weak continuity of the former in the time discrete context is linked to the lack of weak continuity of the latter in the time continuous context. However, we will see that in practive one can actually choose $\sigma = 0$.

3.5 Application to the space of discrete surfaces

Remark: In this section we discuss practical results and numerical approximation techniques presented in [HRWW12] (for discrete geodesics) and [HRS$^+$16] (for discrete splines). For details we refer to these papers.

In the following we will consider families of discrete surfaces which are pairwise in dense correspondence (*cf*. Def. 2.5). Actually, dense correspondence defines an equivalence relation, *i.e.* we consider a fixed equivalence class. This means, all discrete surfaces are based on the same sets of indices $\mathcal{V}$ and $\mathcal{F}$, respectively.

Definition 3.20 (Shape space of discrete surfaces)**.** Given some representative discrete *reference surface* $\mathbf{S}$, the shape space of discrete surfaces $\mathcal{M}[\mathbf{S}]$ is given by the equivalence class of $\mathbf{S}$ where the equivalence relation is given by dense correspondence (as defined in Def. 2.5).

Remark: In the following we assume that we are dealing with an arbitrary but fixed equivalence class $\mathcal{M} = \mathcal{M}[\mathbf{S}]$.

There are two important implications of Def.3.20:

1. The definition of a dissimilarity measure on $\mathcal{M}$ is well-defined, since for two given discrete surfaces $\mathbf{S}, \tilde{\mathbf{S}}$ having the same connectivity, a piecewise affine deformation $\Phi : \mathbf{S} \to \tilde{\mathbf{S}}$ is uniquely determined.
2. The discrete space $\mathcal{M}$ can be identified with $\mathbb{R}^{3n}$, where n is the number of nodes.

Combining Def. 2.6 and Def. 3.9 yields the notion of discrete geodesics in the space of discrete surfaces, analogously, combining Def. 2.6 and Def. 3.19 yields the notion of discrete splines in that space.

Spotlight on computational effort. The variational formulation of the discrete geodesics leads to the following necessary optimality conditions:

$$\begin{aligned} & 0 = \partial_{\mathbf{S}_k}\mathbf{E}^K[\mathbf{S}_0,\dots,\mathbf{S}_K]\,, \quad k = 1,\dots,K-1\,, \\ \iff \quad & 0 = \partial_2\mathbf{W}[\mathbf{S}_{k-1},\mathbf{S}_k] + \partial_1\mathbf{W}[\mathbf{S}_k,\mathbf{S}_{k+1}]\,, \quad k = 1,\dots,K-1\,, \end{aligned} \tag{3.20}$$

where $\partial_i\mathbf{W}$ refers to the variation with respect to the ith argument of $\mathbf{W}$. Note that for a functional $F = F[X]$ we make use of the notation

$$0 = \partial_X F[X] \quad :\Leftrightarrow \quad 0 = \frac{d}{dt}(F[X+tV])\Big|_{t=0} \;\forall\, V \in \mathcal{X}\,,$$

where the test directions V live in a suitable test space $\mathcal{X}$, which is simply $\mathcal{X} = \mathbb{R}^{3n}$ if we consider variations of discrete shell energies. To compute discrete geodesics we have to solve the system of nonlinear equations (3.20) simultaneously, where we fix the two end shapes $\mathbf{S}_0$ and $\mathbf{S}_K$. Hence we are dealing with a nonlinear optimization problem in $\mathbb{R}^{3n(K-1)}$. Similarly, the Euler-Lagrange equations of a discrete spline also result in a nonlinear optimization problem in $\mathbb{R}^{3n(K+1-I)}$ along with $K-1$ nonlinear conditions in $\mathbb{R}^{3n}$ to solve for the geodesic midpoints in (3.18).

Efficient approximation. For a discrete shell $\mathbf{S} \in \mathbb{R}^{3N}$ let $\mathcal{N}$, $\mathcal{E}$ and $\mathcal{F}$ denote the set of vertices, edges and triangles, respectively. Let $L[\mathbf{S}] = (l_e[\mathbf{S}])_e \in \mathbb{R}^{|\mathcal{E}|}$ be the vector of edge lengths. If $e = t_1 \cap t_2$ is the common edge of triangles t_1 and t_2 we associate to e the area measure $d_e = \frac{1}{3}(a_{t_1} + a_{t_2})$. The dihedral angle θ_e of e is defined as the angle between the face normals of t_1 and t_2. Finally, let $\Theta[\mathbf{S}] = (\theta_e[\mathbf{S}])_e \in \mathbb{R}^{|\mathcal{E}|}$ denote the vector of dihedral angles. Now we aim at defining a novel discrete dissimilarity measure in the space of discrete shells that is supposed to an approximation of the one defined in Def. 2.6. To this end, we consider the *Discrete Shells* energy introduced in [GHDS03]. This energy consists of the simplified bending energy (2.31) plus a simplification of a membrane energy. In detail, the simplified membrane energy measures deviations in edge lenghts and triangle volume quadratically. However, we will focus on the length term here. To this end we define for two discrete shells $\mathbf{S}$ and $\tilde{\mathbf{S}}$ the simplified discrete dissimilarity measure via

$$\mathbf{W}^{\text{qu}}[\mathbf{S}, \tilde{\mathbf{S}}] = \mathbf{W}^{\text{qu}}_{\text{mem}}[\mathbf{S}, \tilde{\mathbf{S}}] + \eta\, \mathbf{W}^{\text{qu}}_{\text{bend}}[\mathbf{S}, \tilde{\mathbf{S}}] \tag{3.21}$$

with bending weight $\eta > 0$ and

$$\mathbf{W}^{\text{qu}}_{\text{mem}}[\mathbf{S}, \tilde{\mathbf{S}}] = \sum_{e \in \mathcal{E}} \frac{d_e[\mathbf{S}]}{l_e^2[\mathbf{S}]} \Big(l_e[\mathbf{S}] - l_e[\tilde{\mathbf{S}}] \Big)^2, \quad \mathbf{W}^{\text{qu}}_{\text{bend}}[\mathbf{S}, \tilde{\mathbf{S}}] = \sum_{e \in \mathcal{E}} \frac{l_e^2[\mathbf{S}]}{d_e[\mathbf{S}]} \Big(\theta_e[\mathbf{S}] - \theta_e[\tilde{\mathbf{S}}] \Big)^2.$$

In order to remedy the problem of high computational cost, we introduce a change of variables in order to turn the nonlinear optimization problem into a linear one. We heavily build on the two-step approximation scheme proposed in [FB11] for this change of coordinates. Indeed, for a discrete shell $\mathbf{S} \in \mathbb{R}^{3N}$ we consider the vectors of edge lengths $L = L[\mathbf{S}]$ and dihedral angles $\Theta = \Theta[\mathbf{S}]$ as primary degrees of freedom. The key observation is that with these degrees of freedom, the energy in (3.21) become *quadratic* provided that the purple colored terms are *not* part of the optimization. To achieve this, we replace the purple colored terms in (3.21) by quantities computed on a *reference mesh* $\hat{\mathbf{S}}$. Collecting the new primary variables in one variable $\mathbf{Z} = (L, \Theta)$ living in the $L\Theta$ configuration space $\mathcal{L} := \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}^{|\mathcal{E}|} = \mathbb{R}^{2m}$, we get an approximation of (3.21) via

$$\widehat{\mathbf{W}}_{L\Theta}[\mathbf{Z}, \tilde{\mathbf{Z}}] = \widehat{\mathbf{W}}_{L}[L, \tilde{L}] + \eta\, \widehat{\mathbf{W}}_{\Theta}[\Theta, \tilde{\Theta}] = \sum_{e \in \mathcal{E}} \hat{\alpha}_e \Big(l_e - \tilde{l}_e \Big)^2 + \eta \sum_{e \in \mathcal{E}} \hat{\beta}_e \Big(\theta_e - \tilde{\theta}_e \Big)^2 \tag{3.22}$$

where $\hat{\alpha}_e = d_e[\hat{\mathbf{S}}]/l_e^2[\hat{\mathbf{S}}]$ and $\hat{\beta}_e = l_e^2[\hat{\mathbf{S}}]/d_e[\hat{\mathbf{S}}]$ for all $e \in \mathcal{E}$ and a *reference mesh* $\hat{\mathbf{S}}$. Note that the $\hat{\cdot}$ indicates that the functional is now quadratic but dependent on the reference meshes. We refer to this as the $L\Theta$-energy.

Optimization in $L\Theta$-space. The problem of computing a discrete geodesic and a discrete spline, respectively, can now be re-formulated as follows. For instance, we construct a discrete spline curve in the $L\Theta$-space defined as a minimizer of

$$\hat{F}^K_{L\Theta}[\mathbf{Z}_0, \dots, \mathbf{Z}_K] = 4K^3 \sum_{k=1}^{K} \widehat{\mathbf{W}}_{L\Theta}(\mathbf{Z}_k, \tilde{\mathbf{Z}}_k)\,, \tag{3.23}$$

where $(\mathbf{Z}_{k-1}, \tilde{\mathbf{Z}}_k, \mathbf{Z}_{k+1})$ is a geodesic in the $L\Theta$-space, i.e.,

$$\tilde{\mathbf{Z}}_k = \operatorname*{arg\,min}_{\mathbf{Z} \in \mathcal{L}} \Big(\widehat{\mathbf{W}}_{L\Theta}[\mathbf{Z}_{k-1}, \mathbf{Z}] + \widehat{\mathbf{W}}_{L\Theta}[\mathbf{Z}, \mathbf{Z}_{k+1}] \Big)\,.$$

Since $\widehat{\mathbf{W}}_{L\Theta}$ is *quadratic*, one obtains the explicit solution $\tilde{\mathbf{Z}}_k = \frac{1}{2}(\mathbf{Z}_{k-1} + \mathbf{Z}_{k+1})$. This can be inserted into (3.23) so that we end up with an unconstrained optimization problem. Hence a minimizer of $\hat{F}^K_{L\Theta}$ is a (weighted) cubic spline in the linear space $\mathcal{L} = \mathbb{R}^{2m}$.

Notice that there is no *spatial* coupling between any two different edge lengths in a minimizer $(\mathbf{Z}_0, \dots, \mathbf{Z}_K)$ of (3.23), i.e., an edge length l_e^k of the k^{th} pose interacts only with lengths l_e^j of the same edge e and poses $j \neq k$. The same applies for dihedral angles. As a consequence, the Euler-Lagrange equation for $\hat{F}^K_{L\Theta}$ splits into numerous independent $(K+1)$-dimensional linear systems, i.e., one for each edge length and dihedral angle, which can be solved efficiently and in parallel. Moreover, the matrices representing these linear systems are all given by

$$\begin{pmatrix} 1 & -2 & 1 & & & \\ -2 & 3 & -4 & 1 & & \\ 1 & -4 & 6 & -4 & 1 & \\ & 1 & -4 & 6 & -4 & 1 \\ & & & & \ddots & \end{pmatrix} \in \mathbb{R}^{K+1,K+1},$$

which coincides (for interior quantities) to the 2^{nd} order finite difference approximation of 4^{th} derivatives.
Computing discrete geodesics works analogously, here one obtains a discrete geodesic $(\mathbf{Z}_0, \dots, \mathbf{Z}_K)$ in $\mathcal{L}$ for fixed endpoints $\mathbf{Z}_0$ and $\mathbf{Z}_K$ in terms of intermediate points given by $\mathbf{Z}_k = ((K-k)\mathbf{Z}_0 + k\mathbf{Z}_K)/K$, $k = 1, \dots, K-1$.

Mesh reconstruction. Intermediate values for edge lengths and dihedral angles are generally not realizable as a triangle mesh. Indeed, for closed meshes we have $m \approx 3n$ and hence the map that associates a set of edge lengths and dihedral angles to a given mesh cannot be surjective. To this end, we consider a reconstruction in a least squares sense, similar to [FB11]. For given optimal values $\mathbf{Z}_k = (L_k, \Theta_k)$ we define $\mathbf{S}_k$ as the minimizer of the nonlinear mapping

$$\mathbf{S} \mapsto \widehat{\mathbf{W}}_{L\Theta}(\mathbf{Z}[\mathbf{S}], \mathbf{Z}_k)\,, \tag{3.24}$$

where $\widehat{\mathbf{W}}_{L\Theta}$ is defined as in (3.22). We find the minimizer via the Gauss–Newton method (see Section 6 in [FB11] for details, the "target" values are given by $\mathbf{Z}_k$).
The reconstruction can be seen as a projection of the point $\mathbf{Z} \in \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}^{|\mathcal{E}|}$ onto the submanifold which is given by all sets of edge lengths and dihedral angles that are actually realizable as an embedded triangle mesh. Necessary conditions for points to lie in this submanifold are given by the Gauss-Codazzi equations, see, e.g., [WLT12] for a discrete version. Computationally, the reconstruction is the hardest part in the $L\Theta$-space approximation method. Fortunately it can be parallelized since the reconstruction of one mesh does not depend on the reconstruction of any othere mesh.

Remark: For the discrete splines, optimal $\mathbf{Z}$-variables obtained as solutions of linear systems may have negative lengths. This happens rarely in practice and is most easily addressed by setting corresponding edge weights $\hat{\alpha}_e$ in (3.22) to zero.

3.6 Variational time-discretization of geodesic calculus

Let $y : I \to \mathcal{M}$ be a curve and $W : I \to T\mathcal{M}$ a vector field along y. In Def. 3.6 we defined the covariant derivative $\frac{D}{dt}W$ of W along y for $t \in I$ by

$$\frac{D}{dt}W(t) = \dot{W}(t) + \Gamma_{y(t)}(W(t), \dot{y}(t))\,.$$

With a notion of a covariant derivative along a curve we can define a *parallel transport*, which is indeed well-defined due to Thm. 3.3/3.4 in [Lan95, VIII]:

Proposition 3.21 (Parallel transport)**.** *Let $y : I \to \mathcal{M}$ be a curve. A vector field $V : I \to T\mathcal{M}$ along y is called* parallel *if $\frac{D}{dt}V(t) = 0$ for all $t \in I$. For $t_0 \in I$, $V_0 \in T_{y(t_0)}\mathcal{M}$, there is a unique parallel vector field $V : I \to T\mathcal{M}$ with $V(t_0) = V_0$. Furthermore, the map $\mathrm{P}_{y(t_0)\to y(t)} : T_{y(t_0)}\mathcal{M} \to T_{y(t)}\mathcal{M}$, $\mathrm{P}_{y(t_0)\to y(t)}V_0 = V(t)$ is a linear isomorphism.*

For a given vector $V_0 \in T_{y(t_0)}\mathcal{M}$ one can solve $\frac{D}{dt}V(t) = 0$ with $V(t_0) = V_0$ as an ordinary differential equation to perform the (unique) parallel transport of V_0 along the path.

Remark: As mentioned before, a curve $y : I \to \mathcal{M}$ is usually defined to be *geodesic* if it solves the *geodesic equation*, *i.e.*

$$\frac{D}{dt}\dot{y}(t) = 0 \quad \forall t \in I\,.$$

This means, that y transports its own velocity vector parallely.

Finally, we define the exponential map for a general manifold:

Definition 3.22 (Exponential map)**.** Let $y(t) = y(t, p, V) : I \to \mathcal{M}$, $0 \in I$, be the solution of $\frac{D}{\mathrm{d}t}\dot{y}(t) = 0$ for initial data $y(0) = p$ and $\dot{y}(0) = V$. The (geometric) exponential map $\exp_p : T_p\mathcal{M} \to \mathcal{M}$ is defined as $\exp_p(V) = y(1, p, V)$.

To see that this definition is well-defined we refer to Prop. 4.2 in [Lan95, IV]. Obviously, we have the scaling property $y(t, p, V) = \exp_p(tV)$, which implies that $y(1, p, V)$ is well-defined if $\|V\|$ is sufficiently small. From the local uniqueness of short geodesic paths we deduce that there exists $\delta > 0$, such that $\exp_p : B_\delta(0) \to \exp_p(B_\delta(0))$ is a bijection. In this case, we define $U_p = \exp_p(B_\delta(0))$ to be the *normal neighbourhood* of p. Hence the notion of an inverse mapping is locally well-defined:

Definition 3.23 (Logarithm)**.** The inverse operator of the exponential map is called the (geometric) logarithm $\log_p : U_p \to T_p\mathcal{M}$, where U_p denotes the normal neighbourhood of p.

Now we introduce a time-discretization of these geometric objects. Let $p, q \in \mathcal{M}$ such that there is a unique geodesic $y : [0, 1] \to \mathcal{M}$ with $y(0) = p$ and $y(1) = q$. Then, by Def. 3.23, the logarithm of q with respect to p is the initial velocity $\dot{y}(0) \in T_p\mathcal{M}$, *i.e.* $\log_p(q) = \dot{y}(0)$. The initial velocity $\dot{y}(0)$ can be approximated by a difference quotient in time,

$$\dot{y}(0) = \frac{y(\tau) - y(0)}{\tau} + O(\tau)\,.$$

Thus, we obtain

$$\tau \log_p(q) = y(\tau) - y(0) + O(\tau^2)\,.$$

This gives rise to a consistent definition of a time-discrete logarithm (see [RW15]):

Definition 3.24 (Discrete logarithm)**.** Suppose the discrete geodesic $(y_0, \dots, y_K)$ is the unique minimizer of the discrete path energy (3.9) with $y_0 = p$ and $y_K = q$. Then we define the *discrete logarithm* $\left(\frac{1}{K}\mathrm{LOG}\right)_p(q) = y_1 - y_0$. Note that $\frac{1}{K}$ is part of the symbol and not a factor.

We consider the difference $y_1 - y_0$ as a tangent vector at $p = y_0$. In the special case $K = 1$ we have $\left(\frac{1}{1}\mathrm{LOG}\right)_p(q) = q - p$. As in the continuous case, the discrete logarithm can be considered as a representation of the nonlinear variation q of p in the (linear) tangent space of *displacements*[7] on p.

Next we consider the discretization of the exponential map. In the continuous setting, the exponential map $\exp_p$ maps tangent vectors $V \in T_p\mathcal{M}$ onto the end point $y(1)$ of the unique geodesic $(y(t))_{t\in[0,1]}$ with $y(0) = p$ and $\dot{y}(0) = V$. That means, we have $\exp_p(V) = y(1)$ and, via a simple scaling argument, $\exp_p(t_k V) = y(t_k)$, for $k = 0, \dots, K$, where $t_k = k\tau$ and $\tau = K^{-1}$. In he following we translate this into the discrete setup. Let us again consider a discrete geodesic $(y_0, \dots, y_K)$ with $y_0 = p$ and $y_K = q$. Now $V = \left(\frac{1}{K}\mathrm{LOG}\right)_p(q) = y_1 - y_0$ is the discrete logarithm in the tangent space $T_p\mathcal{M}$, we aim at defining a discrete power k exponential map EXP_p^k such that

$$\mathrm{EXP}_p^k(V) = \mathrm{EXP}_p(kV) = y_k\,.$$

This notation is motivated by the observation that $\exp(ks) = \exp^k(s)$ on $\mathbb{R}$ or more general matrix groups. Furthermore, we would like to have the following recursive property, which holds in the continuous setup:

$$y(t_k) = \exp_p(kV) = \exp_{y(t_{k-2})}(2V_{k-1})\,, \quad V_{k-1} := \log_{y(t_{k-2})} y(t_{k-1})\,, \quad k \geq 2\,. \tag{3.25}$$

That means, once we have defined a discrete version EXP_p^2 corresponding to $\exp_p(2\cdot)$, we can use the recursive relation (3.25) to define EXP_p^k for $k \geq 2$ by

$$y_k = \mathrm{EXP}_p^k(V_1) = \mathrm{EXP}_{y_{k-2}}^2(V_{k-1})\,, \quad V_{k-1} = y_{k-1} - y_{k-2}\,, \tag{3.26}$$

for given $y_0 = p$ and $y_1 = y_0 + V_1$, as shown in Fig. 6.

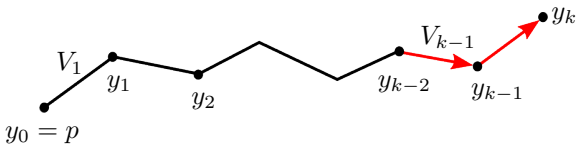


Figure 6: A sketch of the polygonal path associated with the computation of $\mathrm{EXP}_p^k(V_1)$.

[7] Note that these displacements are indeed well-defined, as we assumed that $\mathcal{M}$ embeds into a Banach space.

Note that the discrete analogon of V_{k-1} is exactly $V_{k-1} = (\frac{1}{1}\mathrm{LOG})_{y_{k-2}} y_{k-1} = y_{k-1} - y_{k-2}$.

It remains to define a discrete version EXP_p^2 corresponding to $\exp_p(2\cdot)$. Formally, we have the identity $\frac{1}{2}\log_p\big(\exp_p(2V)\big) = V$, *i.e.* we can define $\mathrm{EXP}^2_{y_0}(y_1 - y_0)$ as the root of the function

$$z \mapsto (\tfrac{1}{2}\mathrm{LOG})_{y_0}(z) - (y_1 - y_0)$$

for given $y_0, y_1 \in \mathcal{M}$. In fact, we are seeking for a third point $y_2 \in \mathcal{M}$, such that (y_0, y_1, y_2) is a time-discrete geodesic for $K = 2$. Using Def. 3.9, a necessary condition of this is given by

$$0 = \partial_2 \mathcal{W}[y_0, y_1](\psi) + \partial_1 \mathcal{W}[y_1, y_2](\psi) \quad \forall \psi \in \mathbf{V}\,,$$

where $\partial_i \mathcal{W}$ denotes the Gâteaux derivative with respect to the ith argument of $\mathcal{W}$. Hence we define:

Definition 3.25 (Discrete exponential map)**.** For given points $y_0, y_1 \in \mathcal{M}$, $V_1 = y_1 - y_0$, we define $\mathrm{EXP}^2_{y_0}(V_1)$ as the solution of

$$\partial_2 \mathcal{W}[y_0, y_1](\psi) + \partial_1 \mathcal{W}[y_1, y](\psi) = 0 \quad \forall \psi \in \mathbf{V}\,,$$

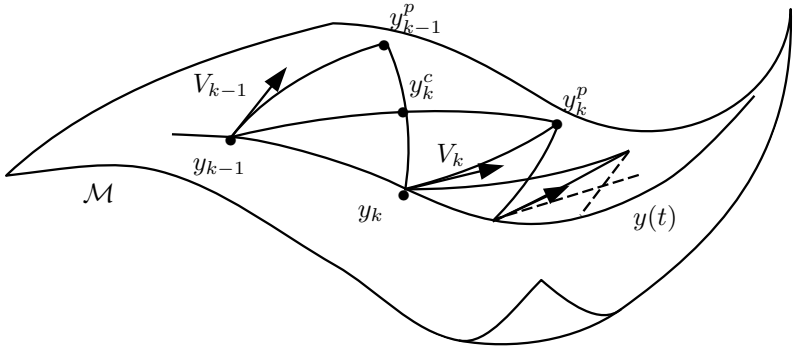
and hence $\mathrm{EXP}^k_{y_0}(V_1) = \mathrm{EXP}^2_{y_{k-2}}(V_{k-1})$ for $V_{k-1} = y_{k-1} - y_{k-2}$ and $k \geq 2$.

It is straightforward to verify that $\mathrm{EXP}_p^K = (\frac{1}{K}\mathrm{LOG})_p^{-1}$ as long as the discrete logarithm is invertible. In fact, the Euler–Lagrange equations for $(y_0, \ldots, y_K)$ being a discrete geodesic with fixed end points y_0 and y_K are given by the $K - 1$ nonlinear equations

$$0 = \partial_2 \mathcal{W}[y_{k-1}, y_k](\psi) + \partial_1 \mathcal{W}[y_k, y_{k+1}](\psi) \quad \forall \psi \in \mathbf{V}\,, \quad k = 1, \ldots, K-1\,, \tag{3.27}$$

which have to be solved *simultaneously*. On the other hand, if we compute $\mathrm{EXP}^k_{y_0}(y_1 - y_0)$ for given $y_0, y_1 \in \mathcal{M}$ and $k = 2, \ldots, K$, we get exactly the same system (3.27). However, in this case the system can be solved *sequentially*.

Finally, we introduce a time-discrete notion of parallel transport along a discrete path as proposed in [RW15]. In the continuous setting, given a path $y : [0, 1] \to \mathcal{M}$ and a vector $V_0 \in T_{y(0)}\mathcal{M}$, parallel transport $\mathrm{P}_{y(0)\to y(\tau)} V_0$ of V_0 along the path y is defined as the solution of the initial value problem $\frac{D}{dt} V(t) = 0$ for $t \in [0, \tau]$ and $V(0) = V_0$. There is a well-known first-order approximation of parallel transport called Schild's ladder (*cf*. [EPS72, KMN00]), which is based on the construction of a sequence of so-called *geodesic parallelograms*; this method has been used *e.g*. by Lorenzi *et al*. [LAP11] to perform parallel transport of deformations along time series of images (see also [PL11]). We once more use the notation $y_k = y(t_k)$, $t_k = k\tau$, for samples of the path $y : [0, 1] \to \mathcal{M}$. Given a tangent vector $V_{k-1} \in T_{y_{k-1}}\mathcal{M}$, the approximation $V_k \in T_{y_k}\mathcal{M}$ of the parallely transported vector $\mathrm{P}_{y_{k-1}\to y_k} V_{k-1}$ via a geodesic parallelogram is illustrated in Fig. 7.
The scheme in Fig. 7 can be easily transferred to the time-discrete setup by replacing y by a discrete path $(y_0, \ldots, y_K)$ and the geodesics that define the geodesic parallelogram by time-discrete geodesics, *e.g*. of length 3. Conceptually, we will again replace tangent or velocity vectors V by displacements ζ of points.
Remark: Let $p_0, p_1, p_2 \in \mathcal{M}$. We define $\hat{p} = \hat{p}(p_0, p_1, p_2)$ such that (p_0, p^c, p_2) and $(p_1, p^c, \hat{p})$ are discrete geodesics for some $p^c \in \mathcal{M}$. Then $(p_0, p_1, p_2, \hat{p})$ defines a *discrete geodesic parallelogram*, p^c is refered to as center point of the parallelogram.



$$y^p_{k-1} = \exp_{y_{k-1}}(V_{k-1})$$

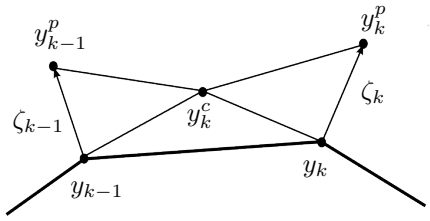
$$y^c_k = \exp_{y^p_{k-1}}\left(\tfrac{1}{2}\log_{y^p_{k-1}}(y_k)\right)$$

$$y^p_k = \exp_{y_{k-1}}\left(2\log_{y_{k-1}}(y^c_k)\right)$$

$$V_k = \log_{y_k}(y^p_k)$$

Figure 7: A sketch of the parallel transport of $V_{k-1} \in T_{y_{k-1}}\mathcal{M}$ from y_{k-1} to y_k along y via Schild's ladder. Here, y^c_k is the midpoint of the two diagonals of the geodesic parallogramm, *i.e.* (y^p_{k-1}, y^c_k, y_k) and (y_{k-1}, y^c_k, y^p_k), which are both geodesic curves.

Definition 3.26 (Discrete parallel transport)**.** Let $(y_0, \dots, y_K)$ be a discrete path in $\mathring{\mathcal{M}}$ with $y_k - y_{k-1}$ sufficiently small for $k = 1, \dots, K$ and ζ_0 a sufficiently small displacement of y_0, given as $y^p_0 = y_0 + \zeta_0$. Then the *discrete parallel transport* of ζ_0 along $(y_0, \dots, y_K)$ is defined for $k = 1, \dots, K$ via the iteration



$$y^c_k = y^p_{k-1} + \left((\tfrac{1}{2}\mathrm{LOG})_{y^p_{k-1}}(y_k)\right),$$

$$y^p_k = \mathrm{EXP}^2_{y_{k-1}}\left(y^c_k - y_{k-1}\right),$$

where $\zeta_k = y^p_k - y_k$ is the transported displacement at y_k. We define

$$\mathbf{P}_{y_K,\dots,y_0}(y^p_0 - y_0) = y^p_K - y_K\,.$$

The notation is chosen such that $\mathbf{P}_{y_K,\dots,y_0}\mathbf{P}_{\tilde{y}_K,\dots,\tilde{y}_0} = \mathbf{P}_{y_K,\dots,y_0,\tilde{y}_K,\dots,\tilde{y}_0}$.

In the kth step of the discrete parallel transport the Euler–Lagrange equations to determine y^c_k and $y^p_k = y_k + \zeta_k$ for given $y^p_{k-1} = y_{k-1} + \zeta_{k-1}$ and discrete path $(y_0, \dots, y_K)$ are

$$\begin{aligned}
\mathcal{W}_{,2}[y^p_{k-1}, y^c_k](\psi) + \mathcal{W}_{,1}[y^c_k, y_k](\psi) &= 0 \quad \forall \psi \in \mathbf{V}\,,\\
\mathcal{W}_{,2}[y_{k-1}, y^c_k](\psi) + \mathcal{W}_{,1}[y^c_k, y^p_k](\psi) &= 0 \quad \forall \psi \in \mathbf{V}\,.
\end{aligned}$$

If $\mathcal{W}$ is symmetric, these conditions are the same as the Euler–Lagrange equations for inverse parallel transport, so that $\mathbf{P}^{-1}_{y_K,\dots,y_0} = \mathbf{P}_{y_0,\dots,y_K}$. However, if $\mathcal{W}$ is not symmetric this is not true in general.

References

[Bal77] John M. Ball. Convexity conditions and existence theorems in nonlinear elasiticity. *Arch. Ration. Mech. Anal.*, 63:337–403, 1977.

[Bär00] Christian Bär. *Elementare Differentialgeometrie*. Walter de Gruyter, 2000.

[Bar15] Sören Bartels. *Numerical Methods for Nonlinear Partial Differential Equations*, volume 47 of *Springer Series in Computational Mathematics*. Springer, 2015.

[BER15] Benjamin Berkels, Alexander Effland, and Martin Rumpf. Time discrete geodesic paths in the space of images. *SIAM J. Imaging Sci.*, 8(3):1457–1488, 2015.

[BKP+10] Mario Botsch, Leif Kobbelt, Mark Pauly, Pierre Alliez, and Bruno Lévy. *Polygon Mesh Processing*. A K Peters, 2010.

[BMF03] Robert Bridson, Sebastian Marino, and Ronald Fedkiw. Simulation of clothing with folds and wrinkles. In *Proc. of ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, 2003.

[Bra02] Andrea Braides. Γ-*convergence for beginners*, volume 22 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2002.

[Bra07] Dietrich Braess. *Finite elements*. Cambridge University Press, Cambridge, third edition, 2007. Theory, fast solvers, and applications in elasticity theory, Translated from the German by Larry L. Schumaker.

[BWH+06] Miklos Bergou, Max Wardetzky, David Harmon, Denis Zorin, and Eitan Grinspun. A quadratic bending model for inextensible surfaces. In *Proc. of Eurographics Symposium on Geometry Processing*, 2006.

[Cia88] Philippe G. Ciarlet. *Mathematical elasticity. Vol. I*, volume 20 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 1988. Three-dimensional elasticity.

[Cia00] Philippe G. Ciarlet. *Mathematical elasticity. Vol. III*, volume 29 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 2000. Theory of shells.

[Cia05] Philippe G. Ciarlet. *An Introduction to Differential Geometry with Applications to Elasticity*. Springer, Dordrecht, 2005.

[CLR04] Ulrich Clarenz, Nathan Litke, and Martin Rumpf. Axioms and variational problems in surface parameterization. *Comput. Aided Geom. Design*, 21(8):727–749, 2004.

[CM08] Philippe G. Ciarlet and Christinel Mardare. An introduction to shell theory. In *Differential geometry: theory and applications*, volume 9 of *Ser. Contemp. Appl. Math. CAM*, pages 94–184. Higher Ed. Press, Beijing, 2008.

[CSM03] David Cohen-Steiner and Jean-Marie Morvan. Restricted Delaunay triangulations and normal cycle. In *Proc. of Symposium on Computational Geometry*, pages 312–321, 2003.

[dB63] Carl de Boor. Best approximation properties of spline functions of odd degree. *J. Math. Mech.*, 12:747–749, 1963.

[dC76] Manfredo P. do Carmo. *Differential geometry of curves and surfaces*. Prentice Hall, 1976. Translated from the Portuguese.

[dC92] Manfredo P. do Carmo. *Riemannian geometry*. Birkhäuser Boston Inc., 1992. Translated from the second Portuguese edition.

[DGDM83] Ennio De Giorgi and Gianni Dal Maso. Γ-convergence and calculus of variations. In *Mathematical theories of optimization (Genova, 1981)*, volume 979 of *Lecture Notes in Math.*, pages 121–143. Springer, Berlin, 1983.

[DH02] Peter Deuflhard and Andreas Hohmann. *Numerische Mathematik. I*. de Gruyter Lehrbuch. [de Gruyter Textbook]. Walter de Gruyter & Co., Berlin, third edition, 2002. Eine algorithmisch orientierte Einführung. [An algorithmically oriented introduction].

[DHLM05] Mathieu Desbrun, Anil N. Hirani, Melvin Leok, and Jerrold E. Marsden. Discrete exterior calculus, 2005. arXiv:math.DG/0508341 on arxiv.org.

[DKT08] Mathieu Desbrun, Eva Kanso, and Yiying Tong. Discrete differential forms for computational modeling. In *Discrete Differential Geometry*, volume 38 of *Oberwolfach Semin.*, pages 287–324. Birkhäuser, Basel, 2008.

[EPS72] Jürgen Ehlers, Felix A. E. Pirani, and Alfred Schild. The geometry of free fall and light propagation. In *General relativity (papers in honour of J. L. Synge)*, pages 63–84. Clarendon Press, Oxford, 1972.

[FB11] Stefan Fröhlich and Mario Botsch. Example-driven deformations based on discrete shells. *Comput. Graph. Forum*, 30(8):2246–2257, 2011.

[FJM02a] Gero Friesecke, Richard D. James, and Stefan Müller. Rigorous derivation of nonlinear plate theory and geometric rigidity. *C. R. Math. Acad. Sci. Paris*, 334(2):173–178, 2002.

[FJM02b] Gero Friesecke, Richard D. James, and Stefan Müller. A theorem on geometric rigidity and the derivation of nonlinear plate theory from three-dimensional elasticity. *Comm. Pure Appl. Math.*, 55(11):1461–1506, 2002.

[FJMM03] Gero Friesecke, Richard D. James, Maria Giovanna Mora, and Stefan Müller. Derivation of nonlinear bending theory for shells from three-dimensional nonlinear elasticity by Gamma-convergence. *C. R. Math. Acad. Sci. Paris*, 336(8):697–702, 2003.

[GGRZ06] Eitan Grinspun, Yotam Gingold, Jason Reisman, and Denis Zorin. Computing discrete shape operators on general meshes. *Comput. Graph. Forum*, 25(3):547–556, 2006.

[GHDS03] Eitan Grinspun, Anil N. Hirani, Mathieu Desbrun, and Peter Schröder. Discrete shells. In *Proc. of ACM SIGGRAPH/Eurographics Symposium on Computer animation*, pages 62–67, 2003.

[Hee11] Behrend Heeren. Geodätische im Raum von Schalenformen. diploma thesis, 2011.

[Hee16] Behrend Heeren. *Numerical Methods in Shape Spaces and Optimal Branching Patterns*. PhD thesis, University of Bonn, 2016.

[Hir03] Anil N. Hirani. *Discrete Exterior Calculus*. PhD thesis, California Institute of Technology, 2003.

[HP04] Klaus Hildebrandt and Konrad Polthier. Anisotropic filtering of non-linear surface features. *Comput. Graph. Forum*, 23(3):391–400, 2004.

[HRS+14] Behrend Heeren, Martin Rumpf, Peter Schröder, Max Wardetzky, and Benedikt Wirth. Exploring the geometry of the space of shells. *Comput. Graph. Forum*, 33(5):247–256, 2014.

[HRS+16] Behrend Heeren, Martin Rumpf, Peter Schröder, Max Wardetzky, and Benedikt Wirth. Splines in the space of shells. *Comput. Graph. Forum*, 35(5):111–120, 2016.

[HRW17] Behrend Heeren, Martin Rumpf, and Benedikt Wirth. Variational time discretization of Riemannian splines. *IMA J. Numer. Anal.*, 2017. accepted.

[HRWW12] Behrend Heeren, Martin Rumpf, Max Wardetzky, and Benedikt Wirth. Time-discrete geodesics in the space of shells. *Comput. Graph. Forum*, 31(5):1755–1764, 2012.

[Joh65] Fritz John. Estimates for the derivatives of the stresses in a thin shell and interior shell equations. *Comm. Pure Appl. Math.*, 18:235–267, 1965.

[Ken84] David G. Kendall. Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.*, 16(2):81–121, 1984.

[Kir50] Gustav Kirchhoff. Über das Gleichgewicht und die Bewegung einer elastischen Scheibe. *J. Reine Angew. Math.*, 40:51–88, 1850. Journal für die Reine und Angewandte Mathematik. [Crelle's Journal].

[KMN00] Arkady Kheyfets, Warner A. Miller, and Gregory A. Newton. Schild's ladder parallel transport procedure for an arbitrary connection. *Internat. J. Theoret. Phys.*, 39(12):2891–2898, 2000.

[Koi66] Warner Tjardus Koiter. On the nonlinear theory of thin elastic shells. *Nederl. Akad. Wetensch. Proc. Ser. B*, 69:1–54, 1966.

[Lan95] Serge Lang. *Differential and Riemannian Manifolds*, volume 160 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 1995.

[LAP11] Marco Lorenzi, Nicholas Ayache, and Xavier Pennec. Schild's ladder for the parallel transport of deformations in time series of images. In *Proc. of International Conference on Information Processing in Medical Imaging*, volume 6801 of *Lecture Notes in Computer Science*, pages 463–474, 2011.

[LDR95] Hervé Le Dret and Annie Raoult. The nonlinear membrane model as variational limit of nonlinear three-dimensional elasticity. *J. Math. Pures Appl. (9)*, 74(6):549–578, 1995.

[LDR96] Hervé Le Dret and Annie Raoult. The membrane shell model in nonlinear elasticity: a variational asymptotic derivation. *J. Nonlinear Sci.*, 6(1):59–84, 1996.

[LDRS05] Nathan Litke, Mark Droske, Martin Rumpf, and Peter Schröder. An image processing approach to surface matching. In *Proc. of Eurographics Symposium on Geometry Processing*, pages 207–216, 2005.

[Lov88] Augustus Edward Hough Love. The small free vibrations and deformation of a thin elastic shell. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 179:491–546, 1888.

[MDSB02] Mark Meyer, Mathieu Desbrun, Peter Schröder, and Alan H. Barr. Discrete differential-geometry operators for triangulated 2-manifolds. In *Visualization and Mathematics III*, pages 35–57. Springer Berlin Heidelberg, 2002.

[MH94] Jerrold E. Marsden and Thomas J. R. Hughes. *Mathematical foundations of elasticity*. Dover Publications, Inc., New York, 1994. Corrected reprint of the 1983 original.

[MRSS15] Jan Maas, Martin Rumpf, Carola Schönlieb, and Stefan Simon. A generalized model for optimal transport of images including dissipation and density modulation. *ESAIM Math. Model. Numer. Anal.*, 49(6):1745–1769, 2015.

[Per15] Ricardo Perl. *Isogeometric Approximation of Variational Problems for Shells*. PhD thesis, University of Bonn, 2015.

[PL11] Xavier Pennec and Marco Lorenzi. Which parallel transport for the statistical analysis of longitudinal deformations? In *Proc. of Colloque GRETSI*, September 2011.

[RE55] Ronald Samuel Rivlin and Jerald Laverne Ericksen. Stress-deformation relations for isotropic materials. *J. Rational Mech. Anal.*, 4:323–425, 1955.

[Rei95] Ulrich Reif. A unified approach to subdivision algorithms near extraordinary vertices. *Comput. Aided Geom. Design*, 12(2):153–174, 1995.

[RS10] Sergey Repin and Stefan A. Sauter. Estimates of the modeling error for the Kirchhoff-Love plate model. *C. R. Math. Acad. Sci. Paris*, 348(17-18):1039–1043, 2010.

[RW13] Martin Rumpf and Benedikt Wirth. Discrete geodesic calculus in shape space and applications in the space of viscous fluidic objects. *SIAM J. Imaging Sci.*, 6(4):2581–2602, 2013.

[RW15] Martin Rumpf and Benedikt Wirth. Variational time discretization of geodesic calculus. *IMA J. Numer. Anal.*, 35(3):1011–1046, 2015.

[Sch64a] Isaac Jacob Schoenberg. On interpolation by spline functions and its minimal properties. In *On Approximation Theory (Proceedings of Conference in Oberwolfach, 1963)*, pages 109–129. Birkhäuser, Basel, 1964.

[Sch64b] Isaac Jacob Schoenberg. Spline interpolation and best quadrature formulae. *Bull. Amer. Math. Soc.*, 70:143–148, 1964.

[Sch73] Isaac Jacob Schoenberg. *Cardinal spline interpolation*. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1973. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 12.

[Sul08] J. M. Sullivan. Curvatures of smooth and discrete surfaces. *Oberwolfach Seminars: Discrete differential geometry*, 38:175–188, 2008.

[War06] Max Wardetzky. *Discrete Differential Operators on Polyhedral Surfaces - Convergence and Approximation*. PhD thesis, Freie Universität Berlin, 2006.

[WBRS09] Benedikt Wirth, Leah Bar, Martin Rumpf, and Guillermo Sapiro. Geodesics in shape space via variational time discretization. In *Proc. of International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 5681 of *Lecture Notes in Computer Science*, pages 288–302, 2009.

[WBRS11] Benedikt Wirth, Leah Bar, Martin Rumpf, and Guillermo Sapiro. A continuum mechanical approach to geodesics in shape space. *Int. J. Comput. Vis.*, 93(3):293–318, 2011.

[Wir09] Benedikt Wirth. *Variational methods in shape space*. Dissertation, University Bonn, 2009.

[WLT12] Yuanzhen Wang, Beibei Liu, and Yiying Tong. Linear surface reconstruction from discrete fundamental forms on triangle meshes. *Comput. Graph. Forum*, 31(8):2277–2287, 2012.