

Sigma Flows for Image and Data Labeling and Learning Structured Prediction*

Jonas Cassel^{1,4}, Bastian Boll¹, Stefania Petra², Peter Albers^{3,4}, and Christoph Schnörr^{1,4}

¹Institute for Mathematics, Image and Pattern Analysis Group, Heidelberg University

²Institute of Mathematics, Mathematical Imaging Group & CAAPS, University of Augsburg

³Institute for Mathematics, Symplectic Geometry Group, Heidelberg University

⁴Research Station Geometry and Dynamics, Heidelberg University

August 29, 2024

Abstract

This paper introduces the *sigma flow model* for the prediction of structured labelings of data observed on Riemannian manifolds, including Euclidean image domains as special case. The approach combines the *Laplace-Beltrami framework* for image denoising and enhancement, introduced by Sochen, Kimmel and Malladi about 25 years ago, and the *assignment flow approach* introduced and studied by the authors.

The sigma flow arises as Riemannian gradient flow of generalized harmonic energies and thus is governed by a nonlinear geometric PDE which determines a harmonic map from a closed Riemannian domain manifold to a statistical manifold, equipped with the Fisher-Rao metric from information geometry. A specific ingredient of the sigma flow is the mutual dependency of the Riemannian metric of the domain manifold on the evolving state. This makes the approach amenable to machine learning in a specific way, by realizing this dependency through a mapping with compact time-variant parametrization that can be learned from data. Proof of concept experiments demonstrate the expressivity of the sigma flow model and prediction performance.

Structural similarities to transformer network architectures and networks generated by the geometric integration of sigma flows are pointed out, which highlights the connection to deep learning and, conversely, may stimulate the use of geometric design principles for structured prediction in other areas of scientific machine learning.

Keywords: harmonic maps, information geometry, Riemannian gradient flows, Laplace-Beltrami operator, neural ODEs, geometric deep learning.

2020 Mathematics Subject Classification. 53B12, 35R01, 35R02, 62H35, 68U10, 68T05, 68T07.

***Acknowledgements.** This work is funded by the Deutsche Forschungsgemeinschaft (DFG), grant SCHN 457/17-2, within the priority programme SPP 2298: Theoretical Foundations of Deep Learning. This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2181/1 - 390900948 (the Heidelberg STRUCTURES Excellence Cluster).

Contents

1	Introduction	3
1.1	Overview, Motivation	3
1.2	Related Work	4
1.2.1	PDE-Based Image and Multi-Dimensional Data Analysis	4
1.2.2	Harmonic Maps and Geometric Gradient Flows	5
1.2.3	Machine Learning	6
1.3	Contribution and Organization	6
2	Information Geometry and Assignment Flows	7
2.1	Basic Notation	7
2.2	Hessian Geometry of the Probability Simplex	8
2.3	S Flows	9
3	Harmonic Maps and Geometric Diffusion	11
3.1	Riemannian Harmonic Maps	11
3.2	Beltrami Flow and Variants	12
3.2.1	Beltrami Flow	12
3.2.2	Variants: Dynamic Metrics	12
3.3	Anisotropic Image Diffusion	13
4	Sigma Flow Model	15
4.1	Harmonic Energy of Probability Simplex-Valued Mappings	15
4.2	Sigma Flows	18
4.3	Sigma- α Flow	24
4.4	Entropic Potential and Convergence to the Boundary	25
4.5	Comparison to the Continuum Limit of the S Flow	28
4.6	Tangent Space Parametrization	29
5	Experiments and Comparison	32
5.1	Implementation	32
5.2	Comparison to the Discrete S Flow Model	34
5.3	Relation with the Transformer Network Architecture	34
5.4	Synthetic Benchmark: Convergence Behavior	36
5.5	Expressivity of the Sigma Flow Model	36
5.6	Learning the Prediction of Labelings in Terms of a Time-Variant Metric	38
6	Conclusion	42
6.1	Summary	42
6.2	Further Work	42
A	Appendix	44
A.1	Glossary of Symbols	44
A.2	Implementation Details	46

1 Introduction

1.1 Overview, Motivation

Since its beginnings, imaging science has been employing a broad range of mathematical methods [Sch15], including models based on partial differential equations (PDEs), variational methods, probabilistic graphical models and differential geometry. In addition, since more than a decade, machine learning has become an integral part of research in *computer vision* in order to deal with complex real-world scenarios. This trend continues, at a slower rate, in the field of *mathematical imaging* where the quest for explainability in methodological research is more emphasized than in computer vision. Naturally, this synergy between mathematical modeling and machine learning has been elaborated most, so far, in connection with the oldest class of problems of the field, image denoising [EKV23].

The work presented in this paper has been motivated by three lines of research:

- (1) The *Laplace-Beltrami framework* [SKM98] for low-level vision which introduced the mathematical framework of harmonic maps [Jos17, Ch. 9]

$$f: M \rightarrow N \tag{1.1}$$

between two Riemannian manifolds (M, h) and (N, g) , to the field of mathematical imaging and computer vision. f is supposed to minimize the so-called harmonic energy, and the corresponding gradient flow defines a geometric diffusion-type PDE. For the specific case $N = \mathbb{R}$, this boils down to functions $f: M \rightarrow \mathbb{R}$ minimizing the corresponding Dirichlet integral, and specializing to functions $f: S^1 \rightarrow N$ results in the familiar geodesic equations as Euler-Lagrange equation.

In this sense, gradient flows corresponding to the general case (1.1) may be considered as generalized higher-dimensional geodesics. Moreover, by making the Riemannian metric h of the domain manifold M dependent on the evolving state, a broad range of PDE-based models, both established ones and our novel model, may be devised in a systematic way, as shown in the present paper.

- (2) *Assignment flows* [ÅPSS17, Sch20] provide a framework for the analysis of metric data on graphs, including image (feature) data on grid graphs as special case. The basic idea is to adopt products of statistical manifolds, in the sense of information geometry [AN00], as state space (N, g) equipped with the Fisher-Rao metric g , and to model contextual inference by flows which emerge from geometric coupling of individual flows on each factor space. Suitable parametrizations of these couplings are amenable to learning these parameters from data, due to the inherent smoothness of the model. Assignment flows, therefore, may be considered as ‘neural ODEs’ from the viewpoint of machine learning.

The most basic instance of this framework concerns the product manifold of open probability simplices. The resulting assignment flows perform *labeling* of metric data on graphs and may be represented as non-local graph-PDEs [SBS23]. A continuous-domain formulation of a special case of assignment flows, on a flat image domain $M \subset \mathbb{R}^2$, was studied in [SS21].

- (3) The use of geometric methods for representing both domains and data has become an active field of research in *machine learning* as well [BBL⁺17]. This naturally motivates to consider the synergy between classical methods and data-driven machine learning, beyond image denoising.

The *goal of this paper* is to combine these lines of research in order to extend the Laplace-Beltrami framework (1) to an intrinsic approach for metric data *labeling*. This is achieved by choosing the target manifold (N, g) as the relative interior of the probability simplex equipped with Fisher-Rao metric (2). Like the assignment flow approach, the resulting *sigma flow* approach is smooth and amenable to machine learning (3). Figure 1.1 illustrates the leitmotiv of this paper.

‘Sigma flow’ reflects the similarity of our model, from the mathematical point of view, to sigma models of mathematical physics – see Remark 3.2 on page 11 and, e.g., [HKK⁺03, Ch. 8] and [Po198, Sec. 3.7] for a general discussion. This is not surprising since the original Laplace-Beltrami approach [SKM98] has been also motivated by mathematical models of high-energy physics.

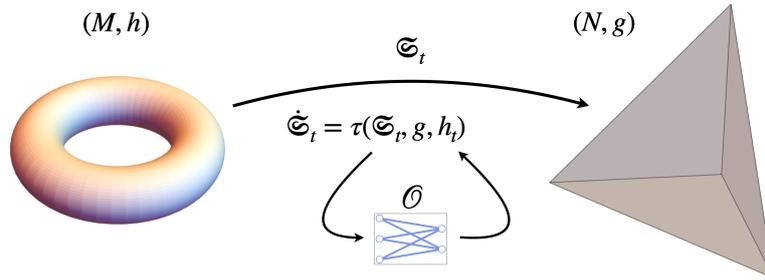


Figure 1.1: **Sigma flow model.** Top-level sketch of the approach introduced in this paper. The approach combines the assignment flow approach (Section 2) and the Laplace-Beltrami framework (Section 3) by choosing the interior of the probability simplex, equipped with the Fisher-Rao metric $g = \mathfrak{g}$, as target manifold N . This naturally extends nonlinear geometric diffusion to metric data labeling through the novel sigma flow model (Section 4). Choosing the metric h of the domain manifold M as a function $h_t = \mathcal{O}(\mathfrak{S}_t)$ of the state \mathfrak{S}_t whose evolution is governed by the sigma flow, enables to represent a range of established models of PDE-based mathematical image analysis (Sections 3.2 and 3.3). In particular, parametrizing the mapping \mathcal{O} by a neural network enables to learn adaptive sigma flows from data. According to the two-stage parametrization $\mathfrak{S}_t \xrightarrow{\mathcal{O}} h_t \xrightarrow{\Delta_{h_t}} \tau(\mathfrak{S}_t, g, h_t) \xrightarrow{\int dt} \mathfrak{S}_t$ of the tension field τ that drives the sigma flow and is generated by the Laplace-Beltrami operator Δ_{h_t} , low-dimensional parameterizations of the mapping \mathcal{O} already achieve strong adaptivity.

Harmonic maps between level surfaces of Hessian domains in the sense of [Shi07], and relative to the α -connections of information geometry [AN00], were studied in [Uoh14]. While our target manifold (N, \mathfrak{g}) is a Hessian manifold, the domain manifold (M, h) , with a metric h learned from data, generally is not. The paper [Uoh14] concludes: “It is an important problem to find applications of non-trivial harmonic maps relative to α -connections.” The approach introduced in the present paper provides such an application using a more general set-up.

1.2 Related Work

No attempt is made here to review the vast literature. We merely point out few prominent works in order to contextualize our paper from the three different viewpoints outlined above.

1.2.1 PDE-Based Image and Multi-Dimensional Data Analysis

PDE-based image analysis has started with the seminal paper [AGLM93], which reports a fundamental study of PDEs whose solutions provide meaningful multiscale transformations of the input data. Here ‘meaningful’ refers to properties like *locality*, *recursivity*, *causality* (for time-variant data) and *invariance* with respect to various transformations. This work still impacts current research. For example, continuous PDE-based formulations of the basic operations of mathematical morphology (dilation, erosion) form the basis for state-of-the-art network architectures that accomplish the equivariant detection of ‘thin structure’ and perceptual grouping in noisy 2D and 3D image data [SPBD23]. For further basic PDE models and the corresponding background, we refer to [Wei98].

Regarding *image denoising*, another line of research based on non-smooth convex functionals and using the total variation (TV) functional as regularizer, has been initiated by [ROF92]. We refer to [BKP10, LRMU15, DMSC16] for advanced generalized TV models and to [CCN15] for a survey. In this context, related variational models for *image labeling* were studied [LS11, CCP12] which constitute convex relaxations of the combinatorial image partitioning problem. From the viewpoint of contemporary research, the inherent non-smoothness of such approaches constitutes a serious obstacle for enhancing model expressivity by parametrization and, in particular, by *learning* parameter values from data.

A powerful class of approaches to *binary image segmentation* in terms of ‘diffusion and threshold’ dynamics was initiated by [MBO94]. A survey of this line of research is provided in [vGGOB14, BF16], where an extension to graphs of the underlying Allen-Cahn equation, as L^2 -gradient of the Ginzburg-Landau functional, is studied. A

drawback of this approach, from our viewpoint, is the lack of a natural formulation of the image and graph data labeling problem with *multiple* labels (non-binary segmentation).

A foundational paper for *geometric PDE*-based image analysis is the *Beltrami flow* introduced in [SKM98], which minimizes an energy functional in terms of the embedding map of a given two-dimensional scalar- or vector-valued image into an Euclidean space. This framework can be used to ‘geometrize’ the aforementioned classical PDE-based approaches, like mean curvature motion induced by the total variation measure, Perona-Malik edge-preserving nonlinear diffusion, etc., in order to achieve also image data *enhancement*, besides image denoising, by representing images as manifolds in Euclidean feature spaces [KMS00]. A *continuous-domain geometric* perspective turned out to be essential also for interpreting graph-Laplacian based denoising schemes in [PC17]. The Laplace-Beltrami framework has been extended to generalized Laplacians on vector bundles in [Bat11] and further generalized to equivariant nonlinear diffusion of vector valued data in [BS14], taking the $SO(3)$ -action on the HSL color space as a case study. Further examples of works which motivate PDEs from various geometric viewpoints (sub-Riemannian geometry, homogeneous spaces) for data denoising, inpainting, enhancement and thin structure detection, include [CFSS16, BCG⁺18, SPSOD21].

Assignment flows [ÅPSS17, Sch20] denote a class of approaches for the analysis of metric data on graphs and for structured prediction. The basic idea, motivated by information geometry [AN00], is to assign to each vertex an elementary statistical manifold as state space equipped with the Fisher-Rao metric and to couple the corresponding Riemannian ascent flows by means of a parametrized affinity function across the graph. Geometric integration of the coupled continuous-time flow [ZSPS20] generates a network with layers indexed by the corresponding discrete points of time, whose parameters are amenable to learning from data by minimizing a suitable loss function [HSPS21]. The choice of a particular statistical manifold as state space depends on the data analysis task. The most basic choice, adopted also in this paper, is the relative interior of the probability simplex for node-wise classification, i.e. data *labeling*. The more expressive case of density matrices as a non-commutative alternative regarding probabilistic models, has been recently introduced and studied in [SCB⁺23].

Major differences of assignment flows to the classes of approaches sketched above include (i) that assignment flows constitute a natural approach to *non-binary* labeling with an arbitrary number of labels and (ii) that integral solutions are obtained after convergence by ‘continuous rounding’, induced by the underlying geometry which couples diffusion and rounding in a *single* process. Stability and convergence to integral solutions was studied in [ZZS22]. Extensions to *unsupervised* and *self-supervised* data labeling were presented in [ZZPS20a, ZZPS20b]. The recent paper [BGAPS24] utilizes randomized assignment flows for the *generative* modeling of high-dimensional joint probability distributions of discrete random variables, via measure transport on the assignment manifold and training by Riemannian flow matching.

Assignment flows can be also represented by a *nonlocal* graph-PDE and shown to comprise a range of related earlier work for data labeling on graphs as special case [SBS23]. The *local* counterpart as PDE on a *continuous* domain was introduced in [SS21], which is the starting point of the present paper.

1.2.2 Harmonic Maps and Geometric Gradient Flows

We focus briefly on the in problem to show the existence and global convergence of gradient flows corresponding to energy functionals which determine harmonic maps and their regularity. This requires to consider more general spaces like $L^2(M, N)$ and the corresponding Sobolev space $W^{1,2}(M, N)$ containing also non-smooth maps f .

For compact Riemannian manifolds M, N , lower-semicontinuity of the harmonic map energy functional with respect to L^2 -convergence is established via Γ -convergence in [Jos17, Ch. 9]. Furthermore, existence is shown assuming that N has nonpositive sectional curvature, employing convexity properties of the energy which can be deduced in this case. The curvature condition and further assumptions that are violated by our models introduced in the present paper, are also adopted for related scenarios studied, e.g., in [Jos97, Nis02, JS09, HJLZ19].

Assuming that M is a compact connected Riemannian manifold with nonvoid boundary and that N is a complete Riemannian manifold, the paper [HKW77] established existence for the corresponding Dirichlet problem, merely assuming a positive upper bound of the sectional curvature of N .

A major relevant line of research is based on the Łojasiewicz-Simon gradient inequality [Sim83, Hua06]. From this angle, the harmonic map problem has been comprehensively studied by [FM19] recently, still assuming that the Riemannian manifold N is closed. The approach requires considerable functional-analytic machinery and a corresponding careful study of the Banach manifold structure of the space of Sobolev $W^{k,p}$ maps.

1.2.3 Machine Learning

The recent paper [CRE⁺21] promotes *graph Beltrami flows* as a proper basis for learning continuous features and evolving the topology of an underlying graph simultaneously. In particular, the authors consider the approach as general enough to overcome a range of limitations of current state-of-the-art deep *graph neural networks (GNNs)*, motivated by the intimate mathematical connection of GNNs to discretized diffusion equations. In fact, there seems to be a trend in machine learning to reconsider concepts like ‘message passing’, ‘attention’ etc. from a mathematical viewpoint and their relation to established concepts (nonlinear, non-local diffusion, continuous-time models, state-dependent inner products, etc.), in order to categorize the great variety of GNN architectures proposed in machine learning during the recent years [HSLG23].

The graph Beltrami flow proposed by [CRE⁺21] considers maps $\mathcal{V} \rightarrow \mathbb{R}^d \times \mathbb{R}^{d'}$ of graph nodes $i \in \mathcal{V}$ to $z_i = (u_i, \alpha x_i)$, $\alpha > 0$, comprising positional encodings u_i computed in a preprocessing step and continuous feature vectors x_i . Defining a discrete gradient operator by finite differences and a discrete divergence operator as adjoint with respect to an inner product, yields the discrete Laplace-Beltrami operator and flow. By constraining the resulting diffusivity, the evolution equation can be written in self-adjoint form and shown to be the gradient flow of a discrete version of the Polyakov action studied in [SKM98].

1.3 Contribution and Organization

In this paper, we adopt the mathematical framework of harmonic maps $f: M \rightarrow N$ in order to extend the assignment flow approach to maps of the form (1.1) and the minimization of a corresponding harmonic energy functional, known as Dirichlet energy in the case of functions $f: M \rightarrow \mathbb{R}$. The target manifold (N, g) will be the interior of the probability simplex $(\overset{\circ}{\Delta}_c, \mathfrak{g})$ equipped with the Fisher-Rao metric \mathfrak{g} . The domain manifold (M, h) with metric h can be any compact Riemannian manifold.

Since our scenario violates basic assumptions made in the literature above (N is open with positive sectional curvature, non-metric affine connection) and generalizes the basic harmonic map problem to sigma models, we leave the problem of existence and global convergence (cf. Section 1.2.2) of the gradient flow for future work and solely focus on *geometric* aspects in this paper.

Throughout the paper, we make the assumptions: (M, h) will be a compact, oriented connected Riemannian manifold without boundary and we consider smooth maps $f \in C^\infty(M, N)$. Specifically, in the case of images, we choose the torus $M = \mathbb{T}^2$ corresponding to the image domain possibly extended by a constant margin, and doubly-periodic boundary conditions, with a metric h induced by data. The compactness assumption ensures the direct application of established results about the spectrum of the Laplace-Beltrami operator, as a basis to devise a Lyapunov functional for the new sigma flow model. This substantiates numerical experiments with N -valued sigma flows on a graph embedded in M , i.e. after a discretization of M .

The main purpose of this paper is to provide, from the viewpoint of geometric modeling and using the framework of harmonic maps, a *continuous-domain* extension $\mathfrak{P}: M \rightarrow N$ of assignment flows with a *learnable time-variant metric* h of the domain manifold M . Since the metric also depends on the evolving state governed by the sigma flow, compact parametrizations enable decent model expressivity.

Our new approach, the *sigma flow* model, is general and applies to labeling tasks of data observed on any compact domain manifold (M, h) . In addition, we consider our result as a mathematical approach to geometric deep learning, contributing a design principle for the generation of neural ODEs by discretizing sigma flows, that accomplish structured data labelings in natural manner using concepts of information and differential geometry. The similarity of our approach to concepts employed in mathematical physics may be of independent interest.

The paper is organized as follows.

Section 2 introduces basic notation, recalls the information geometry of the target manifold $(N, g) = (\overset{\circ}{\Delta}_c, \mathfrak{g})$ and a particular formulation of the assignment flow [SS21] as starting point. The reformulation of the assignment flow approach as nonlinear nonlocal graph PDE [SBS23] characterizes assignment flows as a tool for generating graph-based neural networks for metric data labeling.

Section 3 recalls basic notions related to harmonic maps and the Beltrami flow approach [SKM98]. Variants of this approach are obtained by making the domain metric h dependent on the evolving map. The relation to

nonlinear anisotropic diffusion, in particular, is considered in more detail from this viewpoint.

Section 4 presents our main distribution, the sigma flow model. The general formulation is complemented by concrete implementable expressions using the two basic affine coordinate systems of information geometry. Convergence of the solution are shown under the assumptions stated above about the domain manifold M and the smoothness of mappings f . Finally, we consider the entropic harmonic energy functional which turns the sigma flow model into a proper labeling approach.

Section 5 provides implementation details and few experimental results concerning model expressivity and prediction performance, as proof of concept.

Section 6 Comparisons and structural similarities to the S-flow version of the assignment flow approach (Section 5.2) and to transformer network architectures (Section 5.3), respectively, point out the relevance of the sigma flow model also in a broader context, that we take up and briefly discuss to conclude the paper.

Appendix A.1 lists symbols and their definitions.

Appendix A.2 supplements the specification of implementation details of Section 5.1.

2 Information Geometry and Assignment Flows

We collect few definitions and fix notation which will be used throughout this paper. Appendix A.1 lists symbols and their definitions.

2.1 Basic Notation

Let M be a closed, oriented, connected smooth manifold. For a function $f \in C^\infty(M)$, we use the shorthand

$$\int_M f := \int_M f \omega_h \quad (2.1a)$$

with volume measure ω_h defined by the metric h and locally given as

$$\omega_h = \sqrt{|h|} dx^1 \cdots dx^m =: \sqrt{|h|} dx, \quad (2.1b)$$

where $|h| = \det h$ is the determinant of the metric tensor. We often omit the argument x of functions in integrals, like in (2.1a), to enhance the readability of formulae. For vector bundles E over M , we denote by $\Gamma(E)$ the global sections of E . We furthermore use

$$\Omega^1(M, E) = \Gamma(T^*M \otimes E). \quad (2.2)$$

and $\Omega^1(M) = \Omega^1(M, M \times \mathbb{R})$ for one forms. We set

$$[n] := \{1, 2, \dots, n\}, \quad \forall n \in \mathbb{N} \quad (2.3a)$$

$$[n]_0 := \{0, 1, \dots, n\}, \quad \forall n \in \mathbb{N}. \quad (2.3b)$$

Greek indices denote coordinates for M and roman indices coordinates on the specific target manifold $N = \overset{\circ}{\Delta}_c$ (Section 2.2). In this case, the general Riemannian metric g of $(N, g) = (\overset{\circ}{\Delta}_c, \mathfrak{g})$ is denoted by \mathfrak{g} . Local coordinates on M are denoted by x^μ , $\mu \in [m]$ with coordinate derivative operators

$$\partial_\mu := \frac{\partial}{\partial x^\mu}, \quad \mu \in [m]. \quad (2.4)$$

We denote the number of categories (classes, labels) by $c \in \mathbb{N}$ and set

$$\mathfrak{c} := c - 1. \quad (2.5)$$

The canonical (natural, exponential) local coordinates on $\overset{\circ}{\Delta}_c$ are denoted by θ^i , $i \in [c]$ with coordinate derivative operators

$$\mathfrak{d}_i := \frac{\partial}{\partial \theta^i}, \quad i \in [c]. \quad (2.6)$$

The identity matrix is denoted by

$$\mathbb{I} = \mathbb{I}_n = (\delta_{ij})_{i,j \in [n]}, \quad (2.7)$$

with the Kronecker symbol δ_{ij} , and with the dimension n indicated as subscript whenever the dimension may not be clear from the context. Angular brackets $\langle \cdot, \cdot \rangle$ are generically used for denoting inner products, with the symbol g of the metric as subscript in the case of a Riemannian metric $\langle \cdot, \cdot \rangle_g$. The Einstein summation convention is employed throughout this paper. For two functions $f, g \in C^\infty(M)$, we write

$$\langle df, dg \rangle_h : x \mapsto h^{\mu\nu}(x) \partial_\mu f(x) \partial_\nu g(x) \quad (2.8)$$

for the natural pairing of one-forms induced by h .

2.2 Hessian Geometry of the Probability Simplex

This section defines few basic concepts and notation related to the geometry of the target manifold $(\overset{\circ}{\Delta}_c, g)$. The probability simplex of categorical distributions is denoted by

$$\Delta_c = \left\{ p = (p_0, \dots, p_c) \in \mathbb{R}^c : p_i \geq 0, \sum_{i \in [c]_0} p_i = 1 \right\}. \quad (2.9)$$

Its relative interior is a smooth manifold

$$\overset{\circ}{\Delta}_c = \left\{ p = (p_0, \dots, p_c) \in \mathbb{R}^c : p_i > 0, \sum_{i \in [c]_0} p_i = 1 \right\}. \quad (2.10)$$

Each distribution $p \in \overset{\circ}{\Delta}_c$ has full support and governs a discrete $[c]$ -valued random variable Y with $p_i = \Pr(Y = i)$, $i \in [c]$ and $p_0 = \Pr(Y = c)$. The manifold (2.10) is covered by the two single coordinate charts χ_e and χ_m given by

$$\chi_e : \overset{\circ}{\Delta}_c \rightarrow \mathbb{R}^c, \quad p = (p_0, \dots, p_c) \mapsto \theta = \left(\log \frac{p_1}{p_0}, \dots, \log \frac{p_c}{p_0} \right), \quad (2.11a)$$

$$\chi_m : \overset{\circ}{\Delta}_c \rightarrow \left\{ p \in \mathbb{R}^c : p_i > 0, \sum_{i \in [c]} p_i < 1 \right\}, \quad p = (p_0, \dots, p_c) \mapsto (p_1, \dots, p_c). \quad (2.11b)$$

Denoting the corresponding coordinate functions by

$$\theta^i : \overset{\circ}{\Delta}_c \rightarrow \mathbb{R}, \quad p_i : \overset{\circ}{\Delta}_c \rightarrow \mathbb{R}, \quad i \in [c], \quad (2.12a)$$

one has

$$\theta^i(p) = \log \frac{p_i}{1 - \sum_{j \in [c]} p_j}, \quad p_i(\theta) = \frac{e^{\theta^i}}{1 + \sum_{j \in [c]} e^{\theta^j}}, \quad i \in [c]. \quad (2.12b)$$

The use of sub- and superscripts here is intentional: if θ^i , $i \in [c]$ transform contravariantly, then p_i , $i \in [c]$ transform covariantly. Consequently, $p_i d\theta^i$ is a differential one-form, which turns out to be exact

$$d\psi = p_i d\theta^i, \quad (2.13)$$

with the potential given by the log-Laplace transform (partition function)

$$\psi : \overset{\circ}{\Delta}_c \rightarrow \mathbb{R}, \quad \psi(\theta) = \log \left(1 + \sum_{j \in [c]} e^{\theta^j} \right). \quad (2.14)$$

The Legendre transform yields as conjugate potential

$$\varphi := \psi^* \quad (2.15)$$

the negative entropy

$$\varphi : \overset{\circ}{\Delta}_c \rightarrow \mathbb{R}, \quad \varphi(\theta) = \sum_{j \in [c]} e^{\theta^j - \psi(\theta)} \theta^j - \psi(\theta) = p_j \theta^j - \psi(\theta). \quad (2.16)$$

Basic relations include

$$\mathfrak{d}_i \psi = p_i = e^{\theta^i - \psi}, \quad i \in [c], \quad (2.17a)$$

$$\mathfrak{d}_i \varphi = \mathfrak{g}_{ij} \theta^j, \quad i \in [c], \quad (2.17b)$$

with the metric tensor of the Fisher-Rao metric

$$\mathfrak{g}_{ij} = \mathfrak{d}_i \mathfrak{d}_j \psi = \mathfrak{d}_j p_i = \delta_{ij} p_i - p_i p_j, \quad i, j \in [c]. \quad (2.17c)$$

The Christoffel symbols of the Levi-Civita (metric, Riemannian) connection with respect to the θ coordinates read

$$\Gamma_{jk}^i = \frac{1}{2} \mathfrak{g}^{il} \mathfrak{d}_l \mathfrak{d}_j \mathfrak{d}_k \psi = \frac{1}{2} (\delta_j^i \delta_{jk} - \delta_j^i p_k - \delta_k^i p_j), \quad (2.18)$$

where $\delta_j^i \equiv \delta_{ij}, \forall i, j$. The Christoffel symbols of α -connection in θ -coordinates are then given by [AN00]

$${}^\alpha \Gamma_{jk}^i = \frac{1-\alpha}{2} {}^{-1} \Gamma_{jk}^i + \frac{1+\alpha}{2} {}^1 \Gamma_{jk}^i \quad (2.19a)$$

$$= \frac{1-\alpha}{2} {}^{-1} \Gamma_{jk}^i = (1-\alpha) \Gamma_{jk}^i, \quad i, j, k \in [c], \quad (2.19b)$$

with

$$\frac{1}{2} {}^{-1} \Gamma_{jk}^i = \Gamma_{jk}^i, \quad i, j, k \in [c]. \quad (2.19c)$$

2.3 S Flows

As briefly reported in Section 1.2.1, assignment flows provide a framework for labeling metric data observed on graphs, at every node, utilizing the geometric structure of the simplex $\overset{\circ}{\Delta}_c$ defined above. We confine ourselves to a particular parametrization of assignment flows, called *S Flows* [SS21, Section 3.2].

Let $\mathcal{G}_\Omega = (\mathcal{V}, \mathcal{E}, \Omega)$ be an undirected weighted graph with $\mathcal{V} = [N]$, $N \in \mathbb{N}$ and non-negative weight matrix

$$\Omega \in \mathbb{R}^{N \times N}, \quad \Omega_{ab} \geq 0, \quad \forall a, b \in [N], \quad (2.20)$$

which is symmetric and supported on the edges

$$\Omega_{ab} = \Omega_{ba} \quad \text{and} \quad \Omega_{ab} = 0 \quad \text{if} \quad (a, b) \notin \mathcal{E}. \quad (2.21)$$

The S flow is a dynamical system evolving on the *assignment manifold*

$$(\mathcal{W}_c^N, \mathfrak{g}), \quad \mathcal{W}_c^N := \underbrace{\overset{\circ}{\Delta}_c \times \cdots \times \overset{\circ}{\Delta}_c}_{N \text{ factors}} \quad (2.22)$$

with the Fisher-Rao product metric \mathfrak{g} defined factorwise by (2.17c). Elements S of the assignment manifold \mathcal{W}_c^N are conveniently represented by *assignment matrices* $S \in \mathbb{R}^{N \times c}$ with rows $S_a \in \overset{\circ}{\Delta}_c, a \in [N]$.

The S flow is the Riemannian gradient descent flow corresponding to the objective function

$$\mathcal{J} : \mathcal{W}_c^N \rightarrow \mathbb{R}, \quad \mathcal{J}(S) := -\frac{1}{4} \sum_{a, b \in [N]} \Omega_{ab} \|S_a - S_b\|^2 + \frac{1}{2} \sum_{a \in [N]} \|S_a\|^2 \quad (2.23)$$

given by the equation

$$\dot{S} = -\text{grad } \mathcal{J}(S), \quad S(0) = P. \quad (2.24)$$

The Riemannian gradient can be specified explicitly with some more notation. For a given assignment matrix S we define the *replicator tensor* $R_S \in \mathbb{R}^{N \times c \times c}$ with entries

$$(R_S)_a = \text{Diag}(S_a) - S_a \otimes S_a \in \mathbb{R}^{c \times c}, \quad a \in N. \quad (2.25)$$

The tensor R_S acts on elements of $\mathbb{R}^{N \times c}$ by row-wise matrix multiplication, projecting to the tangent space $T\mathcal{W}_c^N \subset \mathbb{R}^{N \times c}$. For $Y \in \mathbb{R}^{N \times c}$, we write

$$R_S Y \in \mathcal{W}_c^N \quad \text{with rows} \quad (R_S Y)_a = (R_S)_a Y_a. \quad (2.26)$$

Furthermore, we denote by

$$L_\Omega := \Omega - \mathbb{I}_N \in \mathbb{R}^{N \times N} \quad (2.27)$$

the Ω -induced graph-Laplacian, acting on assignment matrices by matrix multiplication. Using this notation, the gradient of J takes the form

$$-\text{grad } J(S) = R_S(L_\Omega S + S). \quad (2.28)$$

The S flow

$$\dot{S} = R_S(L_\Omega S + S) \quad (2.29)$$

is thus as a dynamical system parametrized by the weight matrix Ω . Under mild conditions on Ω [ZZS22] and sufficiently large $T > 0$, $S(T)$ satisfies the entropy constraint

$$H(S_a(T)) < \varepsilon, \quad \forall a \in \mathcal{V}, \quad (2.30)$$

where $H = -\varphi: \mathring{\Delta}_c \rightarrow \mathbb{R}$ denotes the categorical entropy function, see (2.16). This implies that at every node $i \in \mathcal{V}$, the corresponding row $S_a(T) \in \mathring{\Delta}_c$ of the assignment matrix is very close to a unit vector which uniquely assigns the corresponding label to given feature data $f: \mathcal{V} \rightarrow \mathcal{F}$, where \mathcal{F} is any metric space and f is encoded by the initial point P of (2.24).

If the weights Ω are allowed to be adjusted, the S flow ODE (2.29) may be interpreted as a *neural ODE* [CRBD18], where the weights Ω can be learned from data. Among other choices [BCA⁺24], Ω could be parametrized by a deep neural network, as demonstrated e.g. in [BZPS23].

More abstractly, we can think of (2.29) as being parametrized by the Laplace-operator L_Ω rather than by the weights themselves. This perspective is useful regarding generalizations of this formalism, due to the plethora of different structures admitting Laplacian operators, including simplicial complexes [DHLM05, Lim20], meshes [GY02] and manifolds [Jos17]. This general perspective on the S flow is the departure point of the present paper. The goal is to take advantage of the combination of ideas related to data labeling based on assignment flow architectures, with concepts from geometric data processing based on the *manifold hypothesis* [FMN16].

To this end, the paper [SS21] provides a natural starting point for our work, where a continuum limit for the S flow was proposed, replacing the graph \mathcal{G} by an open domain $X \subset \mathbb{R}^d$ for some $d \in \mathbb{N}$. The objective function to be minimized is then replaced by the **S flow energy**

$$E_S : C^\infty(X, \mathring{\Delta}_c) \rightarrow \mathbb{R}, \quad E_S(\mathfrak{P}) = \frac{1}{2} \int_X (\|\partial \mathfrak{P}\|^2 - \|\mathfrak{P}\|^2) dx. \quad (2.31)$$

Numerical optimization of this functional allows to perform data labeling similar to the S flow. In Section 4, we explore how this approach can be generalized when a Riemannian manifold (M, h) is considered instead of a Euclidean domain X , with a metric h depending on data.

3 Harmonic Maps and Geometric Diffusion

3.1 Riemannian Harmonic Maps

Harmonic maps originate in differential geometry when minimizing the energy of functions as defined below. For background and further reading, we refer to [HW08], [Jos17, Ch. 9].

Let

$$(M, h) \quad \text{and} \quad (N, g) \quad (3.1)$$

denote smooth, oriented Riemannian manifolds without boundary of dimensions m and n , respectively. Furthermore, let M be compact. For a function $f \in C^\infty(M, N)$ with coordinate functions $f^i, i = 1, \dots, n$, its differential

$$df \in \Omega^1(M, f^*TN) \quad (3.2)$$

is a section of the vector bundle $T^*M \otimes f^*TN$ of 1-forms with values in the pullback bundle f^*TN over M . The latter is equipped with the metric $g_{ij}(f(x))$ whereas T^*M carries the metric $h^{\mu\nu}(x)$. Thus, denoting the corresponding metric by

$$\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{T^*M \otimes f^*TN}, \quad (3.3)$$

one locally has with the induced norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$,

$$\|df\|^2 \in C^\infty(M), \quad \|df\|^2(x) = h^{\mu\nu}(x) \partial_\mu f^i(x) \partial_\nu f^j(x) g_{ij}(f(x)). \quad (3.4)$$

We call the functional

$$E_H: C^\infty(M, N) \rightarrow \mathbb{R}, \quad E_H(f) = \frac{1}{2} \int_M \|df\|^2 \quad (3.5)$$

the *harmonic energy* of f . Let $f_t: M \times (-\varepsilon, \varepsilon) \rightarrow N$ denote a smooth one-parameter family of variations of $f_0 = f$, then the first variation of the harmonic energy is given by

$$(\delta E_H)(f)[v] := \frac{d}{dt} E_H(f_t) \Big|_{t=0} = - \int_M \langle \tau(f; g, h), v \rangle \omega_h, \quad v = \frac{\partial f_t}{\partial t} \Big|_{t=0}, \quad (3.6)$$

with the **tension field** of f given by

$$\tau(f; g, h) := \text{tr} \nabla df, \quad (3.7)$$

where ∇ denotes the induced connection on $T^*M \otimes f^*TN$. Critical points f of E_H are called *harmonic maps*. The corresponding Euler-Lagrange equations $\tau(f; g, h) = 0$ are more explicitly given by

$$0 = \Delta_h f^i(x) + \Gamma_{jk}^i(f(x)) \partial_\mu f^j(x) \partial_\nu f^k(x) h^{\mu\nu}(x), \quad x \in M, \quad i \in [n], \quad (3.8)$$

with the Christoffel symbols Γ_{jk}^i associated with the metric g on N and the Laplace-Beltrami operator on (M, h) given by

$$\Delta_h f = \frac{1}{\sqrt{|h|}} \partial_\mu (\sqrt{|h|} h^{\mu\nu} \partial_\nu f), \quad f \in C^\infty(M). \quad (3.9)$$

The dependency of the tension field (3.7) on the metrics g and, in particular, on h of (3.1) will be key ingredients of models considered in Sections 3.2.2, 3.3 and 4.

Remark 3.1 (sign convention of Δ_h). The sign convention used here agrees with the one from [SKM98, Eq. (38)] and [Cha84, Section I.1, Eq. (33)], which is opposite to the convention in [Ros97, pp. 18], [Jos17, pp. 496]. Consequently, the operator Δ_h is *negative semi-definite* with a non-positive spectrum.

Remark 3.2 (harmonic maps in theoretical physics). We briefly comment on the role of harmonic maps in theoretical physics. Physical theories describing maps $f: M \rightarrow N$ between manifolds M, N are generally referred to as *sigma models* [HKK⁺03, pp. 146] (the nomenclature is due to historical reasons), where it is often assumed that the metric h is not fixed but dynamical, however. The harmonic energy functional (3.5) is known as the *non-linear sigma model action with target N* in the context of quantum field theory and string theory [Pol98, Sec. 3.7]. A distinguished member of the family of non-linear sigma models is the *Polyakov action* which assumes the special case $N = \mathbb{R}^d$ with the Euclidean metric or Lorentzian pseudo-metric. This case is of great interest in bosonic string theory as laid out in [Pol98, Sec. 1.2] and [Pol81], allowing for tractable quantization.

3.2 Beltrami Flow and Variants

3.2.1 Beltrami Flow

The *Beltrami flow approach* [SKM98] considers images as mappings from surfaces to the RGB color space: A given image array $f \in \mathbb{R}^{N \times 3}$ arises as discretization of a mapping $f : M \rightarrow \mathbb{R}^3$ where, for simplicity, we assume M to be a smooth closed two-dimensional manifold. A basic example are images on a torus $M = \mathbb{T}^2$ with periodic boundary conditions.

Choosing Riemannian metrics

$$(M, h) \quad \text{and} \quad (\mathbb{R}^3, g) \quad (3.10)$$

yields an instance of the harmonic map setting (3.1) and one may consider the harmonic energy $E_H(f)$ of f given by (3.6). The Beltrami flow approach amounts to process f by minimizing E_H and to integrate the corresponding gradient descent equation. Setting for a given $T > 0$

$$M_T := (0, T) \times M \quad (3.11)$$

and for $u \in C^\infty(M_T, \mathbb{R}^3)$ and $t \in (0, T)$ fixed, we write

$$u_t = u(t, \cdot) \in C^\infty(M, \mathbb{R}^3), \quad \text{and} \quad u_0 := \lim_{t \rightarrow 0^+} u(t, \cdot). \quad (3.12)$$

The **basic Beltrami flow** system reads

$$u \in C^\infty(M_T, \mathbb{R}^3) \quad \text{s.t.} \quad \begin{cases} \partial_t u_t = \tau(u_t; g, h), & \forall t \in (0, T), \\ u_0 = f, \end{cases} \quad (3.13)$$

with τ given by (3.7). A common choice is the Euclidean metric

$$g_{ij} = \delta_{ij} \quad (3.14)$$

for the color space. See, e.g., [Res74, Pro16] for color spaces that better conform to human color perception.

3.2.2 Variants: Dynamic Metrics

Variants of the Beltrami flow approach result from coupling the metric h and the function u via a differential equation. Instead of a single metric h , we consider a family of metrics h_t depending on $u_t, t \in (0, T)$. A natural choice is the system

$$u \in C^\infty(M_T, \mathbb{R}^3) \quad \text{s.t.} \quad \begin{cases} \partial_t u_t = \tau(u_t; g, h_t), & \forall t > 0, \\ h_t = u_t^* \delta, & \forall t > 0, \\ u_0 = f, \end{cases} \quad (3.15)$$

with h_t is determined by pulling back the Euclidean metric (3.14) via u_t . A particular case concerns mappings

$$f : M \rightarrow \mathbb{R}^3, \quad f(x) = f(x_1, x_2) = (x_1, x_2, k(x_1, x_2)) \quad (3.16)$$

defined as graph of a function $k : M \rightarrow \mathbb{R}$. Then u_t solving (3.15) defines a family of surfaces governed by the **mean curvature flow** equation

$$\partial_t u_t = 2\mathcal{H}(u_t)\mathcal{N}(u_t), \quad (3.17)$$

where $\mathcal{H}(u_t)$ is the Gaussian mean curvature and $\mathcal{N}(u_t)$ is the unit normal to the surface defined by u_t . For a derivation of this equation, see [SKM98, Sec. 4.3]. We refer, e.g., to [Wei98, SPSOD21] for further reading, to [MBC15] for connections to local adaptive filtering, to [Gar13] for connections to other areas of applied mathematics, and to [Pol98, Sec. 1.2 and 3.7] for relations to theoretical physics.

3.3 Anisotropic Image Diffusion

The anisotropic diffusion approach to image processing promoted by Weickert [Wei98] adopts a somewhat complementary viewpoint. We briefly elucidate differences to, and common aspects with, the Beltrami flow approach.

For the basic case of a gray value image function $f: \mathbb{R}^2 \supset \Omega \rightarrow \mathbb{R}$, the system of **anisotropic diffusion equations** reads

$$u \in C^\infty(\Omega_T) \quad \text{s.t.} \quad \begin{cases} \partial_t u_t = \operatorname{div}(V_t \nabla u_t), & \forall 0 < t < T, \\ V_t = \mathcal{O}(u_t), & \forall t > 0, \\ u_0 = f, \end{cases} \quad (3.18)$$

with partial differential operator \mathcal{O} and a matrix-valued *diffusion tensor* $V: \Omega_T \rightarrow \mathbb{R}^{2 \times 2}$ satisfying a uniform positive definiteness constraint $\mathcal{O}(f) \succeq C(\mathcal{O})\mathbb{I}$, for all $f \in C^\infty(\Omega)$ and a positive constant $C(\mathcal{O})$. The class of operators \mathcal{O} considered in [Wei98] have the form

$$(\mathcal{O}(f))(x) = D(K_\rho * (\nabla f_\sigma \otimes \nabla f_\sigma))(x) \quad (3.19a)$$

$$= D(K_\rho * (f_\sigma^* \delta))(x), \quad f_\sigma := K_\sigma * f, \quad \rho, \sigma > 0, \quad (3.19b)$$

where $D \in C^\infty(\mathbb{R}^{2 \times 2}, \mathbb{R}^{2 \times 2})$ preserves symmetry and positive definiteness of the matrix argument, $*$ denotes spatial convolution and K_ρ, K_σ are lowpass (typically: Gaussian) filter kernels at scales ρ and σ , respectively.

The approach (3.18) has a more narrow scope in that possible manifold structures on Ω are ignored. On the other hand, in view of the second equation of (3.18) governing V_t , Equation (3.19b) generalizes the role of h_t in (3.15).

In order to bring the anisotropic diffusion approach closer to the Beltrami flow approach, we introduce an additional positive warp factor $\kappa: \Omega_T \rightarrow \mathbb{R}_{>}$ and define the **warped anisotropic diffusion (WAD)** system reads

$$u_t \in C^\infty(\Omega_T) \quad \text{s.t.} \quad \begin{cases} \partial_t u = \kappa_t \operatorname{div}(V_t \nabla u_t), & \forall t > 0, \\ V_t = \mathcal{O}(u_t), & \forall t > 0, \\ u_0 = f. \end{cases} \quad (3.20)$$

If we replace the Euclidean domain Ω by a domain manifold M , then the Beltrami flow becomes a special case of the WAD with

$$V_t = h_t^{-1} \sqrt{|h_t|} \quad \text{and} \quad \kappa_t = \frac{1}{\sqrt{|h_t|}}, \quad (3.21)$$

assuming the matrices $h_t \succ 0$ are symmetric and positive definite. This characterizes the Beltrami flow as WAD with a diffusion tensor that has a unit determinant, and it enables to exploit established numerical methods for anisotropic diffusion after choosing a coordinate system on M .

Alternatively, we may generalize the Beltrami flow system to produce WAD systems with more general diffusion tensors. For a C^1 function $B: \mathbb{R} \rightarrow \mathbb{R}$ with $B' > 0$, consider the generalized harmonic energy

$$E_B: C^\infty(M, N) \rightarrow \mathbb{R}, \quad E_B(f) := \frac{1}{2} \int_M B(\|df\|^2). \quad (3.22)$$

Calculating the functional derivative yields

$$(\delta E_B)(f)[v] = - \int_M \langle \tau_B(f; g, h), v \rangle, \quad v = \frac{\partial f_t}{\partial t} \Big|_{t=0}, \quad (3.23)$$

where $\tau(f; g, h) \in \Gamma(f^*TN)$ is locally given by

$$\tau_B(f; g, h)^i = \frac{1}{\sqrt{|h|}} \partial_\mu \left(B'(\|df\|^2) \sqrt{|h|} h^{\mu\nu} \partial_\nu f^i \right) + B'(\|df\|^2) \Gamma_{jk}^i \partial_\mu f^j \partial_\nu f^k h^{\mu\nu}, \quad i \in [n]. \quad (3.24)$$

Choosing specifically $(N, g) = (\mathbb{R}^n, \delta)$, we obtain as a special case of (3.24)

$$\tau_B(f; g, h)^i = \frac{1}{\sqrt{|h|}} \partial_\mu \left(B'(\|df\|^2) h^{\mu\nu} \sqrt{|h|} \partial_\nu f^i \right), \quad i \in [n]. \quad (3.25)$$

The corresponding gradient descent system

$$u \in C^\infty(M_T) \quad \text{s.t.} \quad \begin{cases} \partial_t u_t = \tau_B(u_t; g, h_t), & \forall t \in (0, T), \\ h_t = \mathcal{O}(u_t), & \forall t \in (0, T), \\ u_0 = f. \end{cases} \quad (3.26)$$

Further restricting M to an open domain Ω yields again a system of the form (3.20), but with

$$\det V_t = B'(\|du_t\|^2). \quad (3.27)$$

Prominent special cases include **Perona-Malik denoising** [PM90, Kic08] with

$$h_t = \delta, \quad B'(x) = e^{-x} \quad \text{or} \quad B'(x) = \frac{1}{1+x} \quad (3.28)$$

and **total variation denoising** [ROF92, Cha04] with

$$h_t = \delta, \quad B(x) = 2\sqrt{x}. \quad (3.29)$$

This demonstrates the versatility of the Beltrami flow approach and its variants for representing a range of established methods of PDE-based image processing.

We conclude this section by pointing out two more useful properties of the Beltrami flow approach.

Reparametrization invariance. The harmonic energy E_H (3.5) is invariant under coordinate transformations, i.e. under reparametrizations of the domain manifold M . This provides a basis for coordinate-invariant discretization of the PDE system $\partial_t u_t = \tau(u_t)$. We refer, e.g., to [GY02, War08] for related work.

Conformal invariance. Assume $m = \dim M = 2$. A *conformal transformation* is a rescaling

$$h \rightarrow \lambda h, \quad \lambda: M \rightarrow \mathbb{R}_> \quad (3.30)$$

of the metric h on M with respect to some positive function λ . Writing more explicitly $E_H(f) = E_H(f; g, h)$ for the harmonic energy of the map $f: M \rightarrow N$ between the Riemannian manifolds (M, h) and (N, g) , *conformal invariance* of E_H means

$$E_H(f; g, h) = E_H(f; g, \lambda h), \quad (3.31)$$

which in terms of the functional derivative translates to

$$\delta E_H(f; g, \lambda h) = \frac{1}{\lambda} \delta E_H(f; g, h). \quad (3.32)$$

The consequence for the associated diffusion process is

$$\partial_t u_t = \tau(u_t; g, h) = \lambda \tau(u_t; g, \lambda h), \quad (3.33)$$

which concerns discretization. Setting $\partial_t u \approx \frac{u_{t+\eta} - u_t}{\eta}$ with step size $0 < \eta \ll 1$, one has

$$\frac{u_{t+\eta} - u_t}{\eta} \approx \lambda \tau(u_t; g, \lambda h) \quad \Leftrightarrow \quad u_{t+\eta} \approx u_t + \eta \lambda \tau(u_t; g, \lambda h). \quad (3.34)$$

This shows that the time scale η used for discretization is entangled with the scale λ of the metric h . Since λ is a function varying over the domain M , this also introduces a *spatially resolved* time scale for discretization.

4 Sigma Flow Model

This section presents the main contribution of the paper, the *sigma flow model* for labeling metric data on a smooth compact, oriented closed manifold M equipped with a Riemannian metric h . This is achieved by combining the Beltrami flow and the assignment flow frameworks. Regarding image segmentation, the sigma flow model differs from the methodology presented in [SKM98] in that it works for multiple classes and is an inherently geometric approach to data *labeling*.

Section 4.1 details the Beltrami flow approach for the specific choice

$$(N, g) = (\mathring{\Delta}_c, \mathfrak{g}) \quad (4.1)$$

as target manifold equipped with the Fisher-Rao metric and simplex-valued mappings

$$\mathfrak{P} : M \rightarrow \mathring{\Delta}_c. \quad (4.2)$$

Section 4.2 introduces the sigma flow model and shows that it constitutes a proper geometric diffusion approach. The extension of the sigma flow model from the metric connection to the α -family of connections from information geometry is worked out in Section 4.3. Finally, by additionally taking into account an entropic potential in Section 4.4, the sigma flow model becomes a proper labeling approach.

This version of the novel sigma flow model for data labeling bears resemblance to basic models of mathematical physics (cf. Remarks 3.2 and 4.17) and constitutes the natural geometric extension of the continuous-domain formulation of the assignment flow approach presented by [SS21].

4.1 Harmonic Energy of Probability Simplex-Valued Mappings

Using the notation of Section 3.1, we consider the harmonic energy

$$E_H(\mathfrak{P}) = \frac{1}{2} \int_M \|\mathrm{d}\mathfrak{P}\|^2. \quad (4.3)$$

The corresponding Beltrami flow is generated by the system

$$\mathfrak{S} \in C^\infty(M_T, \mathring{\Delta}_c) \quad \text{s.t.} \quad \begin{cases} \partial_t \mathfrak{S}_t = \tau(\mathfrak{S}_t; \mathfrak{g}, h_t), & \forall t > 0, \\ h_t = \mathcal{O}(\mathfrak{S}_t), & \forall t > 0, \\ \mathfrak{S}_0 = \mathfrak{P}, \end{cases} \quad (4.4)$$

with a differential operator \mathcal{O} specified later and initial condition $\mathfrak{P} \in C^\infty(M, \mathring{\Delta}_c)$. We use both local coordinates

$$\theta^1, \dots, \theta^c \quad \text{and} \quad p_1, \dots, p_c \quad (4.5)$$

of $\mathring{\Delta}_c$ introduced in Section 2.2. The corresponding coordinate expressions of \mathfrak{S} are denoted by

$$\vartheta^1, \dots, \vartheta^c \quad \text{and} \quad \mathfrak{p}_1, \dots, \mathfrak{p}_c, \quad (4.6)$$

respectively.

Proposition 4.1 (harmonic energy on $\mathring{\Delta}_c$). *For a smooth map $\mathfrak{P} : M \rightarrow \mathring{\Delta}_c$ with local θ -coordinate functions ϑ^i , $i \in [c]$ and p -coordinate functions \mathfrak{p}_i , $i \in [c]$, the harmonic energy (4.7) evaluates to*

$$E_H(\mathfrak{P}) = \frac{1}{2} \int_M \mathfrak{g}_{ij} \langle \mathrm{d}\vartheta^i, \mathrm{d}\vartheta^j \rangle_h = \frac{1}{2} \int_M \langle \mathrm{d}\mathfrak{p}_i, \mathrm{d}\vartheta^i \rangle_h, \quad (4.7)$$

with respect to a Riemannian metric h on M .

Proof. By (2.17c), one has $\mathfrak{g}_{ij} = \mathfrak{d}_j \mathfrak{p}_i$ and hence by the chain rule

$$\partial_\mu \mathfrak{p}_i = \mathfrak{d}_j \mathfrak{p}_i \partial_\mu \vartheta^j = \mathfrak{g}_{ij} \partial_\mu \vartheta^j. \quad (4.8)$$

Thus, taking into account the symmetry of \mathfrak{g}_{ij} , (3.5) yields

$$E_H(\mathfrak{P}) = \frac{1}{2} \int_M h^{\mu\nu} \partial_\mu \vartheta^i \partial_\nu \vartheta^j \mathfrak{g}_{ij} = \frac{1}{2} \int_M h^{\mu\nu} \partial_\mu \vartheta^j \partial_\nu \vartheta^i \mathfrak{g}_{ij} = \frac{1}{2} \int_M h^{\mu\nu} \partial_\mu \mathfrak{p}_i \partial_\nu \vartheta^i \quad (4.9a)$$

$$= \frac{1}{2} \int_M \langle d\mathfrak{p}_i, d\vartheta^i \rangle_h. \quad (4.9b)$$

□

We compute the key ingredient of the sigma flow, the tension field τ from Eq. (3.7), in local coordinates.

Proposition 4.2 (tension field in coordinates). *The tension field of $\mathfrak{P} : M \rightarrow \mathring{\Delta}_c$ with θ coordinate functions $\vartheta^i, i \in [c]$ is given locally by*

$$\tau(\mathfrak{P}; \mathfrak{g}, h)^i = \Delta_h \vartheta^i + \frac{1}{2} \langle d(\vartheta^i - \Psi), d(\vartheta^i - \Psi) \rangle_h - \frac{1}{2} \langle d\Psi, d\Psi \rangle_h, \quad i \in [c], \quad (4.10)$$

where $\Psi = \psi \circ \mathfrak{P} : M \rightarrow \mathbb{R}$ from (2.14).

Proof. (3.7) reads

$$\tau(\mathfrak{P}; \mathfrak{g}, h)^i = \Delta_h \vartheta^i + \Gamma_{jk}^i \partial_\mu \vartheta^j \partial_\nu \vartheta^k h^{\mu\nu}. \quad (4.11)$$

Using the relation $\mathfrak{p}_i = \mathfrak{d}_i \psi$ by (2.17a) and hence by the chain rule

$$\partial_\mu \Psi = \mathfrak{d}_i \psi \partial_\mu \vartheta^i = \mathfrak{p}_i \partial_\mu \vartheta^i, \quad (4.12)$$

substitution of the Christoffel symbols (2.18) yields

$$\tau(\mathfrak{P}; \mathfrak{g}, h)^i = \Delta_h \vartheta^i + \frac{1}{2} (\delta_j^i \delta_{jk} - \delta_j^i \mathfrak{p}_k - \delta_k^i \mathfrak{p}_j) \partial_\mu \vartheta^j \partial_\nu \vartheta^k h^{\mu\nu} \quad (4.13a)$$

$$= \Delta_h \vartheta^i + \frac{1}{2} (\partial_\mu \vartheta^i \partial_\nu \vartheta^i - \mathfrak{p}_k \partial_\mu \vartheta^i \partial_\nu \vartheta^k - \mathfrak{p}_j \partial_\mu \vartheta^j \partial_\nu \vartheta^i) h^{\mu\nu} \quad (4.13b)$$

$$= \Delta_h \vartheta^i + \frac{1}{2} (\partial_\mu \vartheta^i \partial_\nu \vartheta^i - \partial_\mu \vartheta^i \partial_\nu \Psi - \partial_\mu \Psi \partial_\nu \vartheta^i) h^{\mu\nu} \quad (4.13c)$$

$$= \Delta_h \vartheta^i + \frac{1}{2} (\langle d\vartheta^i, d\vartheta^i \rangle_h - 2 \langle d\vartheta^i, d\Psi \rangle_h) \quad (4.13d)$$

which is equal to (4.10). □

Since mathematically equivalent expressions may behave differently when they are evaluated numerically, we derive another expression for the tension field in local coordinates.

Proposition 4.3 (alternative form of the tension field). *The tension field of $\mathfrak{P} : M \rightarrow \mathring{\Delta}_c$ with θ -coordinate functions $\vartheta^i, i \in [c]$ and p -coordinate functions $\mathfrak{p}_i, i \in [c]$ is given locally, with respect to the θ -coordinate system, by*

$$\tau(\mathfrak{P}; \mathfrak{g}, h)^i = \frac{1}{2} (\Delta_h \vartheta^i + \mathfrak{g}^{ij} \Delta_h \mathfrak{p}_j), \quad i \in [c]. \quad (4.14)$$

Proof. In this proof, we use explicitly the volume measure ω_h (2.1), which we denote by $\omega_h = \sqrt{|h|} dx$. We compute the first variation of the harmonic energy given by Proposition 4.1. For any smooth functions $f, g : M \rightarrow$

\mathbb{R} , one has

$$\int_M \langle df, dg \rangle_h \omega_h = \int_M \partial_\mu f \partial_\nu g h^{\mu\nu} \sqrt{|h|} dx \quad (4.15a)$$

$$= - \int_M f \partial_\mu (h^{\mu\nu} \sqrt{|h|} \partial_\nu g) dx \quad (4.15b)$$

$$= - \int_M f \partial_\mu (h^{\mu\nu} \sqrt{|h|} \partial_\nu g) \frac{\sqrt{|h|}}{\sqrt{|h|}} dx \quad (4.15c)$$

$$= - \int_M (f \Delta_h g) \omega_h \quad (4.15d)$$

with

$$\Delta_h g = \frac{1}{\sqrt{|h|}} \partial_\mu (h^{\mu\nu} \sqrt{|h|} \partial_\nu g). \quad (4.16)$$

Thus

$$\frac{d}{dt} E_H(\vartheta + t\eta)|_{t=0} = \frac{d}{dt} \frac{1}{2} \int_M \langle d\mathbf{p}_i(\vartheta + t\eta), d(\vartheta^i + t\eta^i) \rangle_h |_{t=0} \quad (4.17a)$$

$$= \frac{1}{2} \int_M \langle d(\eta^j \mathfrak{d}_j \mathbf{p}_i), d\vartheta^i \rangle_h + \frac{1}{2} \int_M \langle d\mathbf{p}_i, d\eta^i \rangle_h \quad (\text{using } \mathfrak{d}_j \mathbf{p}_i = \mathfrak{g}_{ji} = \mathfrak{g}_{ij}) \quad (4.17b)$$

$$= \frac{1}{2} \int_M \eta^j \langle d\mathfrak{g}_{ij}, d\vartheta^i \rangle_h + \frac{1}{2} \int_M \mathfrak{g}_{ij} \langle d\eta^j, d\vartheta^i \rangle_h - \frac{1}{2} \int_M \eta^i \Delta_h \mathbf{p}_i \quad (4.17c)$$

We rewrite the second integral using partial integration.

$$\int_M \mathfrak{g}_{ij} \langle d\eta^j, d\vartheta^i \rangle_h \stackrel{(2.1)}{=} \int_M \mathfrak{g}_{ij} \langle d\eta^j, d\vartheta^i \rangle_h \omega_h = \int_M \mathfrak{g}_{ij} \partial_\mu \eta^j \partial_\nu \vartheta^i h^{\mu\nu} \sqrt{|h|} dx \quad (4.18a)$$

$$= - \int_M \eta^j \partial_\mu \mathfrak{g}_{ij} \partial_\nu \vartheta^i h^{\mu\nu} \sqrt{|h|} dx - \int_M \eta^j \mathfrak{g}_{ij} \partial_\mu (h^{\mu\nu} \sqrt{|h|} \partial_\nu \vartheta^i) dx \quad (4.18b)$$

$$\stackrel{(4.16)}{=} - \int_M \eta^j \langle d\mathfrak{g}_{ij}, d\vartheta^i \rangle_h \sqrt{|h|} dx - \int_M \eta^j \mathfrak{g}_{ij} \Delta_h \vartheta^i \sqrt{|h|} dx. \quad (4.18c)$$

Substitution in (4.17c) yields

$$\frac{d}{dt} E_H(\vartheta + t\eta)|_{t=0} \quad (4.19a)$$

$$= \frac{1}{2} \int_M \eta^j \langle d\mathfrak{g}_{ij}, d\vartheta^i \rangle_h - \frac{1}{2} \int_M \eta^j \langle d\mathfrak{g}_{ij}, d\vartheta^i \rangle_h - \frac{1}{2} \int_M \mathfrak{g}_{ij} \eta^j \Delta_h \vartheta^i - \frac{1}{2} \int_M \eta^i \Delta_h \mathbf{p}_i \quad (4.19b)$$

$$= -\frac{1}{2} \int_M \mathfrak{g}_{ij} \eta^j \Delta_h \vartheta^i - \frac{1}{2} \int_M \eta^j \Delta_h \mathbf{p}_j = -\frac{1}{2} \int_M \mathfrak{g}_{ij} (\Delta_h \vartheta^i + \mathfrak{g}^{ij} \Delta_h \mathbf{p}_j) \eta^j \quad (4.19c)$$

$$= - \int_M \mathfrak{g}_{ij} \tau(\mathfrak{P}; \mathfrak{g}, h)^i \eta^j, \quad (4.19d)$$

which proves (4.14). \square

Remark 4.4 (Fisher-Rao metric becomes singular). We note that the Fisher-Rao metric

$$\mathfrak{g}_{ij} = p_i \delta_{ij} - p_i p_j, \quad i, j \in [c] \quad (4.20)$$

converges to a singular matrix along paths approaching the boundary of $\overset{\circ}{\Delta}_c \subset \Delta_c$. As a remedy, in the rest of this section, we use a regularized metric given in the θ -coordinate system by

$$\mathfrak{g}_\varepsilon := \mathfrak{g} + \varepsilon \mathbb{I}, \quad 0 < \varepsilon \ll 1, \quad (4.21)$$

that is bounded from below by $\varepsilon \mathbb{I}$. The derivatives of the metric are preserved, however.

Proposition 4.5 (Christoffel symbols of \mathfrak{g}_ε). *The Christoffel symbols of \mathfrak{g}_ε , denoted by Γ_ε , are given by*

$$(\Gamma_\varepsilon)_{ijk} = \Gamma_{ijk} = \frac{1}{2} \mathfrak{d}_i \mathfrak{d}_j \mathfrak{d}_k \psi \quad (4.22a)$$

and

$$(\Gamma_\varepsilon)_{jk}^i = (\mathfrak{g}_\varepsilon)^{il} \Gamma_{ljk}, \quad i, j, k \in [\mathfrak{c}]. \quad (4.22b)$$

Proof. The relation $(\Gamma_\varepsilon)_{ijk} = \Gamma_{ijk} = \frac{1}{2} \mathfrak{d}_i \mathfrak{d}_j \mathfrak{d}_k \psi$ follows by the definition of the connection and $\mathfrak{d}_i(\mathfrak{g}_\varepsilon)_{jk} = \mathfrak{d}_i \mathfrak{g}_{jk}$:

$$(\Gamma_\varepsilon)_{ijk} = \frac{1}{2} (\mathfrak{d}_k(\mathfrak{g}_\varepsilon)_{ij} + \mathfrak{d}_j(\mathfrak{g}_\varepsilon)_{ki} - \mathfrak{d}_i(\mathfrak{g}_\varepsilon)_{jk}) = \frac{1}{2} (\mathfrak{d}_k \mathfrak{g}_{ij} + \mathfrak{d}_j \mathfrak{g}_{ki} - \mathfrak{d}_i \mathfrak{g}_{jk}) = \Gamma_{ijk}. \quad (4.23)$$

Further noting [AN00, Section 3.3]

$$\mathfrak{d}_i \mathfrak{g}_{jk} = \mathfrak{d}_i \mathfrak{d}_j \mathfrak{d}_k \psi \quad (4.24)$$

is permutation invariant shows $\Gamma_{ijk} = \frac{1}{2} \mathfrak{d}_i \mathfrak{d}_j \mathfrak{d}_k \psi$. The second claim is the defining relation of $(\Gamma_\varepsilon)_{jk}^i$. \square

4.2 Sigma Flows

The following definition introduces our new model, defined as a Beltrami flow with dynamic metric h and target manifold $(\mathring{\Delta}_c, \mathfrak{g}_\varepsilon)$.

Definition 4.6 (sigma flow). Let $\mathfrak{P} \in C^\infty(M, \mathring{\Delta}_c)$ and $T > 0, \varepsilon > 0$ be fixed. The *sigma flow* is the system of PDEs

$$\mathfrak{S} \in C^\infty(M_T, \mathring{\Delta}_c) \quad \text{s.t.} \quad \begin{cases} \partial_t \mathfrak{S}_t = \tau(\mathfrak{S}_t; \mathfrak{g}_\varepsilon, h_t), & \forall t \in (0, T), \\ h_t = \mathcal{O}(\mathfrak{S}_t), & \forall t \in (0, T), \\ \mathfrak{S}_0 = \mathfrak{P}, \end{cases} \quad (\sigma\text{-flow})$$

where

$$\mathcal{O} : C^\infty(M, \mathring{\Delta}_c) \rightarrow \Gamma(T^*M \otimes T^*M) \quad (4.25)$$

maps to the set of positive definite symmetric 2-tensors such that for all $\mathfrak{P} \in C^\infty(M, \mathring{\Delta}_c)$

$$\mathcal{O}(\mathfrak{P}) \succeq C(\mathcal{O})\mathbb{I} \quad \text{for some } C(\mathcal{O}) > 0. \quad (4.26)$$

The condition on \mathcal{O} is called *uniform positive definiteness criterion* in [Wei98]. It is trivially satisfied by choosing a fixed metric h independent of the state \mathfrak{S} .

In the following, we denote the coordinate expressions of \mathfrak{S} by ϑ and \mathfrak{p} respectively. To avoid cluttered formulae, we do not indicate the time dependence $t \mapsto \vartheta_t, t \mapsto \mathfrak{p}_t$.

Proposition 4.7 (sigma flow in coordinates). *Set $\varepsilon = 0$ in (4.21). If ϑ, \mathfrak{p} denote the coordinate expressions of \mathfrak{S} in θ and p coordinates respectively, the first equation of the system (σ -flow) takes the form*

$$\partial_t \vartheta^i = \frac{1}{2} (\Delta_{h_t} \vartheta^i + \mathfrak{g}^{ij} \Delta_{h_t} \mathfrak{p}_j), \quad i \in [\mathfrak{c}], \quad (4.27a)$$

and

$$\partial_t \mathfrak{p}_i = \frac{1}{2} (\mathfrak{g}_{ij} \Delta_{h_t} \vartheta^j + \Delta_{h_t} \mathfrak{p}_i), \quad i \in [\mathfrak{c}]. \quad (4.27b)$$

Proof. The first form follows directly from Proposition 4.3. It yields the second form after applying equation (4.8) with ∂_t in place of ∂_μ . \square

Our next goal is to devise a Lyapunov functional for the sigma flow, after two preparatory Lemmata; see Proposition 4.10 below.

Lemma 4.8 (spectrum of Laplace-Beltrami operator). *For any metric h on M , the Laplace-Beltrami operator Δ_h is diagonalizable. The eigenfunctions of Δ_h , $\phi_n \in C^\infty(M)$, $n \in \mathbb{N}_0$ exist and form an orthonormal Hilbert basis of $L^2(M)$. Furthermore, let $\lambda_n, n \in \mathbb{N}$ denote the eigenvalues of Δ_h , i.e.*

$$\Delta_h \phi_n = \lambda_n \phi_n, \quad n \in \mathbb{N}_0. \quad (4.28)$$

Then $\lambda_0 = 0$ and $\lambda_i < 0$ for all $i > 0$.

Proof. See [Cha84, Thm.1]. □

Lemma 4.9 (upper/lower uniform boundedness). *The mapping $\mathcal{B} : \mathring{\Delta}_c \rightarrow \mathbb{R}^{c \times c}$ defined in the θ -coordinate system by*

$$\mathcal{B}_{ij}(\theta) = \mathfrak{g}_{ij}(\theta) + \frac{1}{2} \mathfrak{d}_k \mathfrak{g}_{ij}(\theta) \theta^k, \quad i, j \in [c] \quad (4.29)$$

maps into the set of symmetric matrices and admits the bounds

$$\forall \theta, \quad c_2 \mathbb{I} \succeq \mathcal{B}(\theta) \succeq c_1 \mathbb{I}, \quad c_1 = -\frac{c^2 - 1}{2e}, \quad c_2 = \frac{1}{2} \left(1 + \frac{c^2 - 1}{e} \right), \quad (4.30)$$

where $c_1, c_2 > 0$ only depend on $\dim \mathring{\Delta}_c = c = c - 1$.

Proof. In the proof we suppress the dependence on θ for all quantities. Eq. (2.17c) yields

$$\mathfrak{d}_k \mathfrak{g}_{ij} = \mathfrak{d}_k \mathfrak{d}_i \mathfrak{d}_j \psi \quad (4.31)$$

and implies the symmetry of \mathcal{B} given by (4.29). In the context of Hessian geometry, this relation is referred to as the *Codazzi equation* [Shi07, Prop. 2.1]. To establish the convexity bounds of \mathcal{B} , we compute first its entries. Introducing the notation

$$\Theta := (0, \theta^1, \dots, \theta^c) \quad \text{and} \quad \gamma := \max_{i \in [c]} \Theta^i, \quad (4.32)$$

we define

$$\tilde{\theta}^i := \theta^i - \gamma \leq 0 \quad \text{such that} \quad \theta^i = \gamma + \tilde{\theta}^i. \quad (4.33)$$

Recall from (2.17a) and (2.10) the relations

$$p_i = e^{\theta^i - \psi}, \quad i \in [c] \quad \text{and} \quad p_0 := 1 - \sum_{j \in [c]} p_j = e^{-\psi} \quad (4.34)$$

where by (2.14)

$$\psi = \log \exp(\Theta) = \log \left(1 + \sum_{j \in [c]} e^{\theta^j} \right). \quad (4.35)$$

Using $\gamma \geq 0$ and $\tilde{\theta}^i$, $i \in [c]$ as defined by (4.32) and (4.33), we rewrite

$$\psi = \log \left(1 + \sum_{j \in [c]} e^{\theta^j} \right) = \log \left(e^\gamma (e^{-\gamma} + \sum_{j \in [c]} e^{\theta^j - \gamma}) \right) \quad (4.36a)$$

$$= \gamma + \log \left(e^{-\gamma} + \sum_{j \in [c]} e^{\tilde{\theta}^j} \right) \quad (4.36b)$$

and note that the set $\{-\gamma, \tilde{\theta}^1, \dots, \tilde{\theta}^c\}$ contain only non-positive elements and that at least one of them must be 0 by (4.32), (4.33). Thus $e^{\tilde{\theta}^i} \in (0, 1)$, $i \in [c]$ and $e^{-\gamma} \in (0, 1)$, and at least one of them must be 1 such that

$$\log \left(e^{-\gamma} + \sum_{j \in [c]} e^{\tilde{\theta}^j} \right) \in (0, \log c). \quad (4.37)$$

Consequently, we can write

$$\psi = \gamma + \zeta \quad \text{for some } \zeta \in (0, \log c) \quad (4.38)$$

with

$$\zeta \rightarrow 0 \quad \text{as } \gamma \rightarrow \infty, \quad (4.39)$$

since then ψ given by (4.35) approaches the maximal component of the argument vector (4.32). We finally define the function

$$\text{xexp} : \mathbb{R}_{>} \rightarrow \mathbb{R}_{>}, \quad x \mapsto xe^{-x} \in \left(0, \frac{1}{e}\right). \quad (4.40)$$

Now, rewriting the equation (4.29) defining \mathcal{B} in the form

$$\mathcal{B} = \mathfrak{g} + \frac{1}{2}\mathcal{A}, \quad (4.41)$$

we have

$$\mathcal{A}_{ij} = \mathfrak{d}_k \mathfrak{g}_{ij} \theta^k = \theta_i \delta_{ij} - \theta_i p_j - \theta_j p_i, \quad (4.42a)$$

with

$$\theta_i = \mathfrak{g}_{ij} \theta^j \stackrel{(4.33)}{=} \gamma \sum_{j \in [c]} \mathfrak{g}_{ij} + \mathfrak{g}_{ij} \tilde{\theta}^j. \quad (4.42b)$$

Regarding the first term, we compute

$$\gamma \sum_{j \in [c]} \mathfrak{g}_{ij} = \gamma \sum_{j \in [c]} p_i (\delta_{ij} - p_j) = \gamma p_i \left(1 - \sum_{j \in [c]} p_j\right) = \gamma p_i p_0. \quad (4.43)$$

Invoking the relations (4.33), (4.34) and (4.38), we have

$$p_i = e^{\theta^i - \psi(\theta)} = e^{\tilde{\theta}^i} e^{-\zeta} \quad \text{and} \quad p_0 = e^{-\psi(\theta)} = e^{-\gamma - \zeta} \quad (4.44)$$

and thus obtain for the first sum on the right-hand side of (4.42b)

$$\gamma \sum_{j \in [c]} \mathfrak{g}_{ij}(\theta) = \gamma p_i p_0 = \gamma e^{-\gamma} e^{-2\zeta} e^{\tilde{\theta}^i} \stackrel{(4.40)}{=} \text{xexp}(\gamma) e^{-2\zeta} e^{\tilde{\theta}^i} \in \left(0, \frac{1}{e}\right), \quad (4.45)$$

where the bounds follow from the bounds of (4.37), (4.40) and $e^{\tilde{\theta}^i} \in (0, 1)$, $i \in [c]$.

As for the second term on the right-hand side of (4.42b), we have

$$\mathfrak{g}_{ij} \tilde{\theta}^j \stackrel{(4.20)}{=} \tilde{\theta}^j p_i (\delta_{ij} - p_j) \stackrel{(4.44)}{=} \tilde{\theta}^i e^{\tilde{\theta}^i - \zeta} - \sum_{j \in [c]} \tilde{\theta}^j e^{\tilde{\theta}^j - \zeta} e^{\tilde{\theta}^i - \zeta} \quad (4.46a)$$

$$\stackrel{\tilde{\theta}^i \leq 0}{=} -\text{xexp}(|\tilde{\theta}^i|) e^{-\zeta} + \sum_{j \in [c]} \text{xexp}(|\tilde{\theta}^j|) e^{\tilde{\theta}^i} e^{-2\zeta} \in \left(-\frac{1}{e}, \frac{c}{e}\right), \quad (4.46b)$$

Putting together, we obtain for (4.42b)

$$\theta_i = \gamma \sum_j \mathfrak{g}_{ij} + \tilde{\theta}^j \mathfrak{g}_{ij} \in \left(-\frac{1}{e}, \frac{c}{e}\right), \quad c = \mathfrak{c} + 1. \quad (4.47)$$

From $p_i \in (0, 1)$ follows then

$$\mathcal{A}_{ij} \stackrel{(4.42a)}{=} \theta_i \delta_{ij} - \theta_i p_j - \theta_j p_i \in \left(-\frac{c+1}{e}, \frac{c+1}{e}\right) =: (-q, q), \quad q = \frac{c+1}{e} \quad (4.48)$$

which implies $\|\mathcal{A}\|_\infty \leq cq$ and in turn the bounds of the spectral radius of \mathcal{A}

$$\rho(\mathcal{A}) \leq \|\mathcal{A}\|_\infty \leq cq \quad \implies \quad -cq\mathbb{I} \preceq \mathcal{A} \preceq cq\mathbb{I}. \quad (4.49)$$

Returning to (4.41), we infer

$$\forall v \in \mathbb{R}^c, \quad \mathcal{B}_{ij}v^i v^j = \mathfrak{g}_{ij}v^i v^j + \frac{1}{2}\mathcal{A}_{ij}v^i v^j \stackrel{\mathfrak{g} \succeq 0}{\geq} \frac{1}{2}\mathcal{A}_{ij}v^i v^j \geq -\frac{cq}{2}\|v\|^2 \quad (4.50)$$

which gives the lower bound

$$\mathcal{B} \succeq -\frac{cq}{2}\mathbb{I}. \quad (4.51)$$

As for an upper bound, we first note that with $p_i \in (0, 1)$, $i \in [c]$,

$$\mathfrak{g}_{ii} = p_i - p_i^2 = p_i(1 - p_i) \in \left(0, \frac{1}{4}\right), \quad i \in [c], \quad (4.52a)$$

$$\sum_{j \in [c] \setminus \{i\}} |\mathfrak{g}_{ij}| = p_i \sum_{j \in [c] \setminus \{i\}} p_j = p_i(1 - p_0 - p_i) \leq p_i(1 - p_i) \in \left(0, \frac{1}{4}\right), \quad i \in [c] \quad (4.52b)$$

such that by Gerschgorin's circle theorem [HJ13, Thm. 6.1.1]

$$\rho(\mathfrak{g}) \in \left(0, \frac{1}{2}\right). \quad (4.53)$$

Consequently,

$$\mathcal{B}_{ij}v^i v^j \leq \frac{1}{2}\|v\|^2 + \frac{1}{2}\mathcal{A}_{ij}v^i v^j \leq \frac{1}{2}(1 + cq)\|v\|^2, \quad (4.54)$$

that is

$$\mathcal{B} \preceq \frac{1}{2}(1 + cq)\mathbb{I} \quad (4.55)$$

which together with

$$cq = (c-1)\frac{c+1}{e} = \frac{c^2-1}{e} \quad (4.56)$$

concludes the proof. \square

We are now in the position to devise a Lyapunov functional for the sigma flow.

Proposition 4.10 (Lyapunov functional). *Let $\varphi: \mathring{\Delta}_C \rightarrow \mathbb{R}$ be given by (2.16) and $\varepsilon > 0$ be fixed. Then the functional*

$$\Phi: C^\infty(M, \mathring{\Delta}_c) \rightarrow \mathbb{R}_{\geq}, \quad \Phi(\mathfrak{P}) = \int_M \left(\varphi(\mathfrak{P}) + \frac{\varepsilon}{2} \delta_{ij} \vartheta^i \vartheta^j \right) \quad \vartheta = \chi_\varepsilon \circ \mathfrak{P}: M \rightarrow \mathbb{R}^c \quad (4.57)$$

is a Lyapunov functional for the sigma flow (σ -flow) if

$$\varepsilon + c_1 = \beta > 0, \quad (4.58)$$

with c_1 from (4.30).

Proof. Let $\mathfrak{S} \in C^\infty(M_T, \mathring{\Delta}_c)$ solve the sigma flow system (σ -flow) for $T > 0$ fixed and ϑ_t be the θ -coordinate functions of \mathfrak{S}_t and $h_t = \mathcal{O}(\mathfrak{S}_t)$ the time dependent metric. In order to show that Φ is a Lyapunov functional, we show that $t \mapsto \Phi(\mathfrak{S}_t)$ is bounded from below, continuous, differentiable and monotonically decreasing in time.

Due to convexity of the integrand function (φ is the negative entropy), the functional is bounded from below by $\Phi(0) = -\text{vol}(M) \log c$, where $0 \in C^\infty(M, \mathring{\Delta}_c)$ is the 0 function in the θ -coordinate system. As for the continuity at 0, we have

$$|\Phi(\mathfrak{S}_t) - \Phi(\mathfrak{S}_0)| \leq \text{vol}(M) \max_{x \in M} \left(|\varphi(\vartheta_t(x)) - \varphi(\vartheta_0(x))| + |\delta_{ij} \vartheta_t^i(x) \vartheta_t^j(x) - \delta_{ij} \vartheta_0^i(x) \vartheta_0^j(x)| \right) \quad (4.59a)$$

$$\leq \text{vol}(M) \max_{x \in M} \left(|\varphi(\vartheta_t(x)) - \varphi(\vartheta_0(x))| + (\|\vartheta_t(x)\| + \|\vartheta_0(x)\|) \|\|\vartheta_t(x)\| - \|\vartheta_0(x)\|\| \right). \quad (4.59b)$$

Since φ is continuous and M is compact, the right-hand side goes to 0 as $\vartheta_t \rightarrow \vartheta_0$, which shows the continuity of Φ at 0. Further, note that the integrand of $\Phi(\mathfrak{S}_t)$ is continuously differentiable in time. The compactness of M then implies that $\Phi(\mathfrak{S}_t)$ is differentiable for all $t \in (0, T)$.

We do not indicate the time dependence of quantities in the rest of this proof to alleviate notation. We show now that Φ is monotonically decreasing in time.

$$\frac{d}{dt} \Phi(\mathfrak{S}) = \int_M \mathfrak{d}_i \varphi \frac{\partial}{\partial t} \vartheta^i + \varepsilon \delta_{ij} \vartheta^j \frac{\partial}{\partial t} \vartheta^i \quad (4.60a)$$

$$\stackrel{(2.16)}{=} \stackrel{(2.17b)}{=} \int_M (\mathfrak{g}_{ij} + \varepsilon \delta_{ij}) \vartheta^j \frac{\partial}{\partial t} \vartheta^i = \int_M (\mathfrak{g}_\varepsilon)_{ij} \vartheta^j \frac{\partial}{\partial t} \vartheta^i \quad (4.60b)$$

and using $\partial_t \vartheta^i = \tau(\mathfrak{S}; \mathfrak{g}_\varepsilon, h)^i = \Delta_h \vartheta^i + (\Gamma_\varepsilon)_{jk}^i \langle d\vartheta^j, d\vartheta^k \rangle_h$ by (4.11) (with the metric \mathfrak{g}_ε in place of \mathfrak{g})

$$= \int_M (\mathfrak{g}_\varepsilon)_{ij} \vartheta^j \Delta_h \vartheta^i + \int_M (\mathfrak{g}_\varepsilon)_{ij} (\Gamma_\varepsilon)_{kl}^i \langle d\vartheta^k, d\vartheta^l \rangle_h \vartheta^j. \quad (4.60c)$$

By Prop. 4.5 we have $(\mathfrak{g}_\varepsilon)_{ij} (\Gamma_\varepsilon)_{kl}^i = \Gamma_{klj}$ and $\Gamma_{klj} = \Gamma_{jkl}$ for any order of the indices. Thus

$$\frac{d}{dt} \Phi(\mathfrak{S}) = \int_M (\mathfrak{g}_\varepsilon)_{ij} \vartheta^j \Delta_h \vartheta^i + \int_M \Gamma_{ijk} \vartheta^i \langle d\vartheta^j, d\vartheta^k \rangle_h. \quad (4.60d)$$

Regarding the first integral on the right-hand side, we apply partial integration and use again the volume measure $\omega_h = \sqrt{|h|} dx$ explicitly

$$\int_M (\mathfrak{g}_\varepsilon)_{ij} \vartheta^j \Delta_h \vartheta^i \stackrel{(2.1)}{=} \int_M (\mathfrak{g}_\varepsilon)_{ij} \vartheta^i \Delta_h \vartheta^j \sqrt{|h|} dx \stackrel{(3.9)}{=} \int_M (\mathfrak{g}_\varepsilon)_{ij} \vartheta^i \partial_\nu \left(h^{\mu\nu} \sqrt{|h|} \partial_\mu \vartheta^j \right) dx \quad (4.61a)$$

$$= - \int_M \partial_\nu (\mathfrak{g}_\varepsilon)_{ij} \vartheta^i h^{\mu\nu} \partial_\mu \vartheta^j \sqrt{|h|} dx - \int_M (\mathfrak{g}_\varepsilon)_{ij} \partial_\nu \vartheta^i h^{\mu\nu} \partial_\mu \vartheta^j \sqrt{|h|} dx. \quad (4.61b)$$

Taking into account the chain rule $\partial_\nu (\mathfrak{g}_\varepsilon)_{ij} = \partial_\nu \mathfrak{g}_{ij} = \mathfrak{d}_k \mathfrak{g}_{ij} \partial_\nu \vartheta^k$, we obtain

$$\frac{d}{dt} \Phi(\mathfrak{S}) = - \int_M \mathfrak{d}_k \mathfrak{g}_{ij} \langle d\vartheta^j, d\vartheta^k \rangle_h \vartheta^i - \int_M (\mathfrak{g}_\varepsilon)_{ij} \langle d\vartheta^j, d\vartheta^i \rangle_h + \int_M \Gamma_{ijk} \vartheta^i \langle d\vartheta^j, d\vartheta^k \rangle_h \quad (4.62a)$$

and using the relation $\mathfrak{d}_k \mathfrak{g}_{ij} = 2\Gamma_{kij} = 2\Gamma_{ijk}$

$$\frac{d}{dt} \Phi(\mathfrak{S}) = - \int_M (\mathfrak{g}_\varepsilon)_{ij} \langle d\vartheta^j, d\vartheta^i \rangle_h - \frac{1}{2} \int_M \mathfrak{d}_k \mathfrak{g}_{ij} \langle d\vartheta^j, d\vartheta^k \rangle_h \vartheta^i \quad (4.62b)$$

$$= - \int_M \left(\mathfrak{g}_{ij} + \varepsilon \delta_{ij} + \frac{1}{2} \vartheta^k \mathfrak{d}_k \mathfrak{g}_{ij} \right) \langle d\vartheta^i, d\vartheta^j \rangle_h \quad (4.62c)$$

$$\stackrel{(4.29)}{=} - \int_M (\mathcal{B}(\vartheta) + \varepsilon \mathbb{I})_{ij} \langle d\vartheta^i, d\vartheta^j \rangle_h. \quad (4.62d)$$

The integrand has the form $\text{tr}(AB)$ where A, B are symmetric positive semi-definite matrices (ε was chosen so that this is true). Invoking the lower bound $\lambda_{\min}(B) \text{tr} A \leq \text{tr}(AB)$ implied by a trace inequality [MOA11, p. 341, H.1.h] and taking into account (4.58) gives

$$\frac{d}{dt} \Phi(\mathfrak{S}) \leq -\beta \int_M \delta_{ij} \langle d\vartheta^i, d\vartheta^j \rangle_h. \quad (4.63)$$

By virtue of Lemma 4.8 and (4.28), we expand the functions ϑ^i in the orthonormal basis $(\phi_n)_{n \in \mathbb{N}_0}$ provided by the Laplacian Δ_h ,

$$\vartheta^i = \sum_{n \geq 0} a_n^i \phi_n, \quad a_n^i = \int_M \vartheta^i \phi_n \quad (4.64)$$

to obtain

$$\int_M \langle d\vartheta^i, d\vartheta^j \rangle_h \stackrel{(4.15)}{=} - \int_M \vartheta^i \Delta_h \vartheta^j = - \sum_{n_1, n_2 \geq 0} a_{n_1}^i a_{n_2}^j \lambda_{n_2} \underbrace{\int_M \phi_{n_1} \phi_{n_2}}_{=\delta_{n_1 n_2}} \quad (4.65a)$$

$$= - \sum_{n \geq 0} a_n^i a_n^j \lambda_n \stackrel{\lambda_0=0}{=} - \sum_{n \geq 1} a_n^i a_n^j \lambda_n. \quad (4.65b)$$

Returning to (4.63), we thus have

$$\frac{d}{dt} \Phi(\mathfrak{S}) \leq \beta \sum_{n \geq 1} (\delta_{ij} a_n^i a_n^j) \lambda_n = \beta \sum_{n \geq 1} \|a_n\|^2 \lambda_n \quad (4.66a)$$

with

$$a_n = (a_n^1, \dots, a_n^\varepsilon)^\top. \quad (4.66b)$$

The eigenvalues λ_n of the Laplacian Δ_h depend on the state \mathfrak{S} via the coupling $h = \mathcal{O}(\mathfrak{S})$. We can however bound the metric uniformly from below by $C(\mathcal{O})\mathbb{I}$, $c > 0$ due to (4.26). This allows to give a uniform bound on λ_n as follows. For any $n \geq 1$, invoke the identity [Cha84, Eq. (46)]

$$|\lambda_n| = \int_M h^{\mu\nu} \partial_\mu \phi_n \partial_\nu \phi_n \quad (4.67)$$

with $\int_M \phi_n = 0$, which follows from the fact that eigenfunctions of the Laplacian are either constant or have mean 0 and the constant eigenfunctions are associated with $\lambda_0 = 0$. Since we know that $h^{-1} \succeq \frac{1}{C(\mathcal{O})}\mathbb{I}$ we can apply a trace inequality as in (4.63)

$$|\lambda_n| \geq \frac{1}{C(\mathcal{O})} \int_M \delta^{\mu\nu} \partial_\mu \phi_n \partial_\nu \phi_n = \frac{1}{C(\mathcal{O})} \|d\phi_n\|_\delta^2. \quad (4.68)$$

Then, the Poincaré lemma [Jos17, Cor. A.1.1] guarantees the existence of a constant $C(M) > 0$ such that

$$\|\phi_n\|_{L^2(M)} \leq C(M) \|d\phi_n\|_\delta \quad (4.69)$$

which implies

$$|\lambda_n| \geq \frac{1}{C(\mathcal{O})} \|d\phi_n\|_\delta^2 \geq \frac{1}{C(M)^2 C(\mathcal{O})} \|\phi_n\|_{L^2(M)}^2. \quad (4.70)$$

By normalization of the eigenfunctions $\|\phi_n\|_{L^2(M)} = 1$ we we thus obtain $\lambda_n < -\frac{1}{C(M)C(\mathcal{O})}$. From (4.66a), we finally infer

$$\frac{d}{dt} \Phi(\mathfrak{S}) \leq -\frac{\beta}{C(M)^2 C(\mathcal{O})} \sum_{n \geq 1} \|a_n\|^2. \quad (4.71)$$

Because $\|a_n\|$ measures the non-constant part of \mathfrak{S} if $n > 0$, it follows that $\Phi(\mathfrak{S})$ monotonically decays as long as $\sum_{n \geq 1} \|a_n\|^2 > 0$, and $\sum_{n \geq 1} \|a_n\|^2 = 0$ holds if and only if \mathfrak{S} is constant. \square

Remark 4.11 (sigma flow: existence and convergence). Proposition 4.10 shows that the convex functional Φ given by (4.57) is monotonically decreasing as long as \mathfrak{S}_t is not constant. After discretizing the domain manifold M which is required for numerical experiments, this characterizes the sigma flow as proper geometric diffusion process, i.e. $\lim_{t \rightarrow \infty} \mathfrak{S}_t$ is constant. However, to rigorously show existence and global convergence in the general case, a weak set-up with a feasible set of \mathfrak{S}_t containing $C^\infty(M, \mathring{\Delta}_c)$ as dense subspace would have to be considered, as discussed in Section 1.2.2.

Remark 4.12 (harmonic maps into spheres). A similar conclusion could have been drawn along a different line of reasoning, when considering that harmonic maps into the sphere orthant must be constant, as implied by the general theory of harmonic maps into spheres [Sol85]. The only point to note is that $\mathring{\Delta}_c$ is isometric to the positive orthant of a sphere.

4.3 Sigma- α Flow

The sigma flow system (σ -flow) involves the metric connection $\Gamma = {}^0\Gamma$ of the Fisher-Rao metric. In this section, we consider the extension to the family of α -connections ${}^\alpha\Gamma$ from information geometry given by (2.19). As a consequence, the tension field (3.7) given explicitly by (3.8) will take the form (cf. also (4.11))

$${}^\alpha\tau(\mathfrak{S}_t; \mathfrak{g}, h_t)^i = \Delta_{h_t}\vartheta^i + {}^\alpha\Gamma_{jk}^i \langle d\vartheta^j, d\vartheta^k \rangle_{h_t}, \quad i \in [c]. \quad (4.72)$$

Definition 4.13 (sigma- α flow). Let $\mathfrak{P} \in C^\infty(M, \mathring{\Delta}_c)$ and $\alpha \in \mathbb{R}, T > 0$ be given. The sigma- α flow is generated by the tension field (4.72) and the system

$$\mathfrak{S} \in C^\infty(M_T, \mathring{\Delta}_c) \quad \text{s.t.} \quad \begin{cases} \partial_t \mathfrak{S}_t = {}^\alpha\tau(\mathfrak{S}_t; \mathfrak{g}, h_t), & \forall t \in (0, T), \\ h_t = \mathcal{O}(\mathfrak{S}_t), & \forall t \in (0, T), \\ \mathfrak{S}_0 = \mathfrak{P}, \end{cases} \quad (\sigma\text{-}\alpha \text{ flow})$$

where \mathcal{O} satisfies the uniform positive definiteness condition (4.26).

The following proposition generalizes Proposition 4.7 accordingly. It reveals, in particular, that the sigma- α flow combines two linear flows corresponding to the two extreme cases of the α -connections, viz. the case $\alpha = 1$,

$$\partial_t \vartheta^i = \Delta_{h_t} \vartheta^i, \quad i \in [c] \quad (4.73)$$

and the case $\alpha = -1$,

$$\partial_t \mathfrak{p}_i = \Delta_{h_t} \mathfrak{p}_i, \quad i \in [c]. \quad (4.74)$$

Proposition 4.14 (sigma- α flow in coordinates). *The first equation of the system (σ - α flow) is given with respect to the θ coordinates by*

$$\partial_t \vartheta^i = \frac{1+\alpha}{2} \Delta_{h_t} \vartheta^i + \frac{1-\alpha}{2} \mathfrak{g}^{ij} \Delta_{h_t} \mathfrak{p}_j, \quad i \in [c] \quad (4.75)$$

and with respect to the p coordinates by

$$\partial_t \mathfrak{p}_i = \frac{1+\alpha}{2} \mathfrak{g}_{ij} \Delta_{h_t} \vartheta^j + \frac{1-\alpha}{2} \Delta_{h_t} \mathfrak{p}_i, \quad i \in [c]. \quad (4.76)$$

Proof. By (2.19), Eq. (4.72) reads

$${}^\alpha\tau(\mathfrak{S}_t; \mathfrak{g}, h_t)^i = \Delta_{h_t} \vartheta^i + (1-\alpha) \Gamma_{jk}^i \langle d\vartheta^j, d\vartheta^k \rangle_{h_t} \quad (4.77)$$

and we show below equivalence to (4.75). Equation (4.76) then follows directly using

$$\partial_\mu \mathfrak{p}_i = \mathfrak{g}_{ij} \partial_\mu \vartheta^j. \quad (4.78)$$

Using

$$\Gamma_{jk}^i = \frac{1}{2} \mathfrak{g}^{il} \mathfrak{d}_l \mathfrak{g}_{jk} \quad (4.79)$$

we transform the right-hand side of (4.75) into (4.77).

$$\frac{1+\alpha}{2}\Delta_{h_t}\vartheta^i + \frac{1-\alpha}{2}\mathfrak{g}^{ij}\Delta_{h_t}\mathfrak{p}_j \stackrel{(3.9)}{=} \frac{1+\alpha}{2}\Delta_{h_t}\vartheta^i + \frac{1-\alpha}{2}\mathfrak{g}^{ij}\frac{1}{\sqrt{|h_t|}}\partial_\mu(h_t^{\mu\nu}\sqrt{|h_t|}\partial_\nu\mathfrak{p}_j) \quad (4.80a)$$

$$\stackrel{(4.78)}{=} \frac{1+\alpha}{2}\Delta_{h_t}\vartheta^i + \frac{1-\alpha}{2}\mathfrak{g}^{ij}\frac{1}{\sqrt{|h_t|}}\partial_\mu(h_t^{\mu\nu}\sqrt{|h_t|}\mathfrak{g}_{jk}\partial_\nu\vartheta^k) \quad (4.80b)$$

$$= \frac{1+\alpha}{2}\Delta_{h_t}\vartheta^i + \frac{1-\alpha}{2}\underbrace{\mathfrak{g}^{ij}\mathfrak{g}_{jk}}_{\delta_{ik}}\frac{1}{\sqrt{|h_t|}}\partial_\mu(h_t^{\mu\nu}\sqrt{|h_t|}\partial_\nu\vartheta^k) \quad (4.80c)$$

$$+ \frac{1-\alpha}{2}\mathfrak{g}^{ij}\underbrace{\mathfrak{d}_l\mathfrak{g}_{jk}\partial_\mu\vartheta^l}_{\partial_\mu\mathfrak{g}_{jk}}\frac{1}{\sqrt{|h_t|}}h_t^{\mu\nu}\sqrt{|h_t|}\partial_\nu\vartheta^k \quad (4.80d)$$

$$= \left(\frac{1+\alpha}{2} + \frac{1-\alpha}{2}\right)\Delta_{h_t}\vartheta^i + (1-\alpha)\frac{1}{2}\mathfrak{g}^{ij}\mathfrak{d}_l\mathfrak{g}_{jk}\langle d\vartheta^l, d\vartheta^k \rangle_{h_t} \quad (4.80e)$$

$$\stackrel{(4.79)}{=} \Delta_{h_t}\vartheta^j + (1-\alpha)\Gamma_{lk}^i\langle d\vartheta^l, d\vartheta^k \rangle_{h_t}. \quad (4.80f)$$

□

Remark 4.15 (regularized metric \mathfrak{g}_ε). Convergence of these flows to constant solutions under the assumption of Proposition 4.10, and with the reservation concerning the general case expressed as Remark 4.11, can be shown by minor adaption of the arguments. We omit the details but we note that, if the regularized metric \mathfrak{g}_ε given by (4.21) is to be used instead of \mathfrak{g} , this entails the replacements

$$\mathfrak{g} \rightarrow \mathfrak{g}_\varepsilon, \quad \Gamma \rightarrow \Gamma_\varepsilon, \quad \text{and} \quad \varphi(\mathfrak{p}) \rightarrow \varphi(\mathfrak{p}) + \frac{\varepsilon}{2}\delta_{ij}\vartheta^i\vartheta^j, \quad (4.81)$$

with $\mathfrak{g}_\varepsilon, \Gamma_\varepsilon$ defined by (4.21), (4.22).

4.4 Entropic Potential and Convergence to the Boundary

The geometric diffusion equations introduced so far produce constant solutions in the infinite time limit. This is at odds with the goal of achieving a labeling of observed data at every point $x \in M$, that is an assignment of a definite label. We modify the sigma flow system (σ -flow) to achieve such labelings by including a term that drives the flow to the boundary of the target manifold $(\mathring{\Delta}_c, \mathfrak{g})$.

Definition 4.16 (entropic harmonic energy). Let $m^2 > 0$ be fixed. We define the *entropic harmonic energy* as

$$E_\varphi : C^\infty(M, \mathring{\Delta}_c) \rightarrow \mathbb{R}, \quad \mathfrak{P} \mapsto \int_M \left(\frac{1}{2}\|d\mathfrak{P}\|^2 - m^2\varphi(\mathfrak{P}) \right), \quad (4.82)$$

with φ given by (2.16).

As a consequence of including the entropy term, the expressions (4.11) and (4.14) for the tension field in θ coordinates change to

$$\tau_\varphi(\mathfrak{S}_t; \mathfrak{g}, h)^i = \Delta_h\vartheta^i + \Gamma_{jk}^i\langle d\vartheta^j, d\vartheta^k \rangle_h + m^2\vartheta^i \quad (4.83a)$$

$$= \frac{1}{2}(\Delta_h\vartheta^i + \mathfrak{g}^{ij}\Delta_h\mathfrak{p}_i) + m^2\vartheta^i, \quad i \in [c]. \quad (4.83b)$$

See Remark 4.15 for minor modifications if the metric \mathfrak{g} is replaced by the ε -regularized metric \mathfrak{g}_ε .

Remark 4.17 (potentials in physics). The modification of the sigma flow according to Definition 4.16 is reminiscent of adding a potential function to models of physical systems. In the present case, the potential $-\varphi$ is concave, contrary to most of the common cases in physics. However, scenarios where locally concave potentials appear also in physics include the *Higgs potential* [Ham17, Ch. 8] or the *Landau-Ginzburg potential* [Jos17, Ch. 11], where

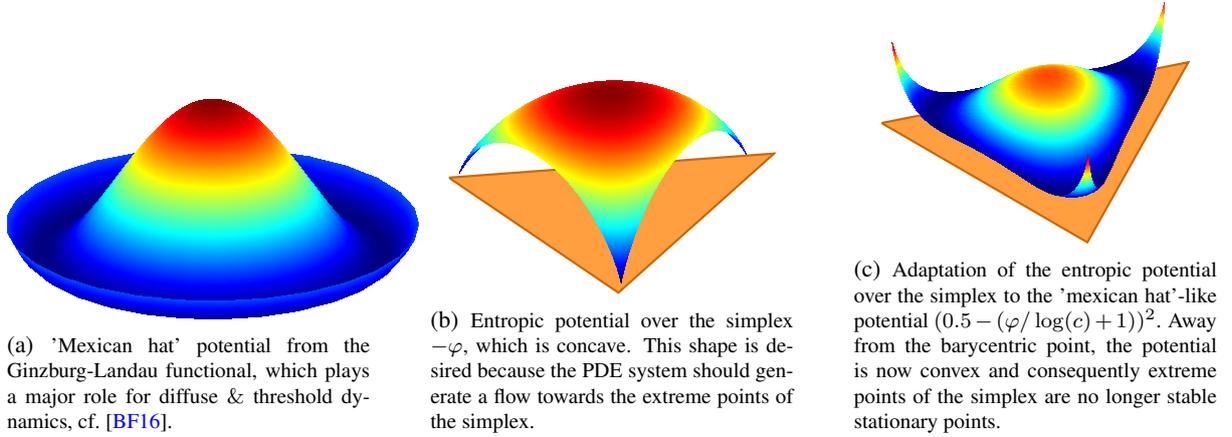


Figure 4.1: Mexican hat potential and entropic potential on the simplex (colored orange).

the potential has the shape of a 'mexican hat' depicted by Figure 4.1. A similar shape could be also achieved in the sigma flow setting by the replacement $-\varphi \rightarrow (0.5 - (\varphi/\log c + 1))^2$ producing a potential that is concave around the origin but convex when approaching the boundary of the simplex.

The shape of the above-mentioned 'mexican hat' potential may remind some readers of the 'mexican hat' shaped convolution masks for edge detection in image data, generated by Laplacian-of-Gaussian operators, which have a long history in early computer vision and as physiological models of simple cells [MH80]. Our class of models introduced in this paper is PDE-based, however, rather than based on convolution followed by thresholding. Specifically, the metric h_t on M is coupled to the evolving state \mathfrak{S} , which may be used for – in comparison to basic convolution and thresholding: sophisticated – edge detection, as in (3.28), for instance. For a detailed study of the connection between models based on PDEs and on convolution, respectively, we refer to [BCM06], and to [BF16] for advanced approaches combining diffusion and threshold dynamics.

Definition 4.18 (entropic sigma flow). Let $\mathfrak{P} \in C^\infty(M, \mathring{\Delta}_c)$, $\varepsilon > 0$, $m^2 > 0$ be fixed. The regularized *entropic sigma flow system* is given by

$$\mathfrak{S} \in C^\infty(M_T, \mathring{\Delta}_c) \quad \text{s.t.} \quad \begin{cases} \partial_t \mathfrak{S}_t = \tau_\varphi(\mathfrak{S}_t; \mathfrak{g}_\varepsilon, h_t), & \forall t \in (0, T), \\ h_t = \mathcal{O}(\mathfrak{S}_t), & \forall t \in (0, T), \\ \mathfrak{S}_0 = \mathfrak{P}, & \end{cases} \quad (\sigma\text{-}\varphi\text{-flow})$$

where \mathcal{O} satisfies the uniform positive definiteness condition (4.26).

Theorem 4.19 (entropic sigma flow: convergence). Assume that $\mathfrak{S} \in C^\infty(M_\infty, \mathring{\Delta}_c)$ is a solution of ($\sigma\text{-}\varphi\text{-flow}$) existing for all time and let $m^2, \varepsilon > 0$ be fixed but arbitrary. Let $t > 0$ be fixed and ϑ denote the θ coordinate expression of \mathfrak{S}_t , and let $\vartheta^i = \sum_{n \geq 0} a_n^i \phi_n$ be the decomposition of ϑ into eigenfunctions of the Laplacian Δ_{h_t} , analogous to (4.64). Define the set of low frequencies as

$$\mathfrak{N} = \{n \in \mathbb{N}_0 : c_2 \lambda_n + \varepsilon(\lambda_n + m^2) > 0\} \quad (4.84)$$

with c_2 from (4.30) and assume

$$a_n^i = 0 \quad \text{for all } n \notin \mathfrak{N}, \quad i \in [c]. \quad (4.85)$$

If condition (4.85) holds for all $t > 0$, then the $L^2(M, \mathring{\Delta}_c)$ -norm of \mathfrak{S}_t is unbounded as a function of time.

It is clear that the set \mathfrak{N} is never empty; it always contains 0. For ε fixed, \mathfrak{N} can be enlarged by increasing m^2 .

Proof. To alleviate notation, we drop the subscript t in this proof for all time dependent quantities. Consider the functional $\Phi : C^\infty(M, \mathring{\Delta}_c) \rightarrow \mathbb{R}$ as in Proposition 4.10. The proof consists in showing that $\Phi(\mathfrak{S})$ is strictly

increasing in time. Hence we compute $\frac{d}{dt}\Phi(\mathfrak{S})$ in the θ -coordinate system, where \mathfrak{S} satisfies the differential equation (σ - φ -flow)

$$\frac{\partial}{\partial t}\mathfrak{S} = \tau_\varphi(\mathfrak{S}; \mathfrak{g}, h), \quad h = \mathcal{O}(\mathfrak{S}) \quad (4.86)$$

which evaluates in θ -coordinates to

$$\frac{\partial}{\partial t}\vartheta^i \stackrel{(4.83)}{=} \Delta_h \vartheta^i + \Gamma_{jk}^i \langle d\vartheta^j, d\vartheta^k \rangle_h + m^2 \vartheta^i. \quad (4.87)$$

Similarly to Eq. (4.60b) one derives

$$\frac{d}{dt}\Phi(\mathfrak{S}) = \int_M (\mathfrak{d}_i \varphi + \varepsilon \delta_{ij} \vartheta^j) \frac{\partial}{\partial t} \vartheta^i = \int_M (\mathfrak{g}_\varepsilon) \vartheta^j \frac{\partial}{\partial t} \vartheta^i \quad (4.88)$$

yielding

$$\frac{d}{dt}\Phi(\mathfrak{S}) = \int_M (\mathfrak{g}_\varepsilon)_{ij} \vartheta^j (\Delta_h \vartheta^i + \Gamma_{jk}^i \langle d\vartheta^j, d\vartheta^k \rangle_h + m^2 \vartheta^i). \quad (4.89)$$

The first two terms can then be treated as in the proof of Prop. 4.10 where the equation

$$\int_M (\mathfrak{g}_\varepsilon) \vartheta^j (\Delta_h \vartheta^i + \Gamma_{jk}^i \langle d\vartheta^j, d\vartheta^k \rangle_h) \stackrel{(4.62d)}{=} \int_M (\mathcal{B} + \varepsilon \mathbb{I})_{ij} \langle d\vartheta^i, \vartheta^j \rangle_h \quad (4.90)$$

is shown with \mathcal{B} defined as in Lemma 4.9. The last expression can be bounded by

$$- \int_M (\mathcal{B} + \varepsilon \mathbb{I})_{ij} \langle d\vartheta^i, d\vartheta^j \rangle_h \geq -(c_2 + \varepsilon) \delta_{ij} \int_M \langle d\vartheta^i, d\vartheta^j \rangle_h \quad (4.91)$$

since $c_2 \mathbb{I}$ is an upper bound to \mathcal{B} by (4.30) and due to a trace inequality argument similar to the one used in Eq. (4.63). The last expression can be further simplified when expressed in terms of the expansion coefficients (4.64)

$$- \int_M \langle d\vartheta^i, d\vartheta^j \rangle_h \stackrel{(4.65)}{=} \sum_{n \geq 0} a_n^i a_n^j \lambda_n \stackrel{(4.85)}{=} \sum_{n \in \mathbb{N}} a_n^i a_n^j \lambda_n. \quad (4.92)$$

Putting together the results from above yields

$$\int_M (\mathfrak{g}_\varepsilon) \vartheta^j (\Delta_h \vartheta^i + \Gamma_{jk}^i \langle d\vartheta^j, d\vartheta^k \rangle_h) \stackrel{(4.90)}{=} - \int_M (\mathcal{B} + \varepsilon \mathbb{I})_{ij} \langle d\vartheta^i, d\vartheta^j \rangle_h \quad (4.93a)$$

$$\stackrel{(4.91)}{\geq} -(c_2 + \varepsilon) \delta_{ij} \int_M \langle d\vartheta^i, d\vartheta^j \rangle_h \quad (4.93b)$$

$$\stackrel{(4.92)}{=} (c_2 + \varepsilon) \delta_{ij} \sum_{n \in \mathbb{N}} a_n^i a_n^j \lambda_n. \quad (4.93c)$$

A for the second term, we invoke the positive definiteness of $\mathfrak{g} \succeq 0$ and obtain a lower bound

$$m^2 \int_M (\mathfrak{g}_\varepsilon)_{ij} \vartheta^j \vartheta^i = m^2 \int_M (\mathfrak{g} + \varepsilon \mathbb{I})_{ij} \vartheta^j \vartheta^i \geq m^2 \varepsilon \delta_{ij} \int_M \vartheta^j \vartheta^i. \quad (4.94)$$

Furthermore, substituting the expansion coefficients gives

$$\int_M \vartheta^j \vartheta^i = \sum_{n \geq 0} a_n^i a_n^j \stackrel{(4.85)}{=} \sum_{n \in \mathbb{N}} a_n^i a_n^j \quad (4.95)$$

and hence the lower bound

$$m^2 \int_M (\mathfrak{g}_\varepsilon)_{ij} \vartheta^j \vartheta^i \geq \varepsilon m^2 \delta_{ij} \sum_{n \in \mathbb{N}} a_n^i a_n^j = \varepsilon m^2 \sum_{n \in \mathbb{N}} \|a_n\|^2. \quad (4.96)$$

Combining the bounds (4.93), (4.96), we thus obtained

$$\frac{d}{dt}\Phi(\mathfrak{S}) = \int (\mathfrak{g}_\varepsilon)\vartheta^j (\Delta_h \vartheta^i + \Gamma_{jk}^i \langle d\vartheta^j, d\vartheta^k \rangle_h) + m^2 \int_M (\mathfrak{g}_\varepsilon)_{ij} \vartheta^j \vartheta^i \quad (4.97a)$$

$$\geq \sum_{n \in \mathbb{N}} ((c_2 + \varepsilon)\lambda_n + \varepsilon m^2) \|a_n\|^2 > 0, \quad (4.97b)$$

because for all $n \in \mathbb{N}$ by (4.84), we have $(c_2 + \varepsilon)\lambda_n + \varepsilon m^2 > 0$. This shows that Φ is monotonously increasing in time. Since φ is bounded from above by 0, however, we conclude that the L^2 -norm of coordinate functions ϑ diverges as a function of time. \square

Remark 4.20 (relevance for convergence in practice). The interpretation of Theorem 4.19 is that the entropic sigma flow converges to the simplex boundary $\partial \mathring{\Delta}_c$ (in the L^2 sense) if only low frequency modes are present in the state \mathfrak{S} . This can be expected to hold due to the diffusion part of the tension field (4.83). Numerical experiments substantiate this result in Section 5.

4.5 Comparison to the Continuum Limit of the S Flow

We compare the entropic harmonic energy functional (4.82) and the S flow functional on the continuous domain, as described by equation (2.31). In order to establish a common ground between the two models, we set $M = X \subset \mathbb{R}^2$ an open domain which is a Riemannian manifold with the Euclidean metric $h = \delta$ induced by \mathbb{R}^2 . The S flow functional E_S considers $\mathfrak{P} : X \rightarrow \mathring{\Delta}_c \subset \mathbb{R}^c$ as a mapping to \mathbb{R}^c and evaluates to

$$E_S(\mathfrak{P}) = \frac{1}{2} \int_X (\|d\mathfrak{P}\|^2 - \|\mathfrak{P}\|^2) d^2x. \quad (4.98)$$

The harmonic energy is reads in θ -coordinates

$$E_\varphi(\mathfrak{P}) = \int_X \left(\frac{1}{2} \mathfrak{g}_{ij} \langle d\vartheta^i, d\vartheta^j \rangle_\delta - \varphi(\mathfrak{P}) \right) d^2x. \quad (4.99)$$

This last expression can be simplified by employing the sphere map defined next.

Definition 4.21 (sphere map [Kas89, AN00]). The sphere map

$$\Lambda : \mathring{\Delta}_c \rightarrow 2S^c; \quad p \mapsto 2\sqrt{p} \quad (4.100)$$

is a smooth isometric immersion that identifies the simplex $\mathring{\Delta}_c$ with the positive orthant of the sphere $2S^c$ of radius 2, equipped with the round metric [ÅPSS17, AJLS17].

Proposition 4.22 (spherical representation). Let $\mathfrak{P} : X \rightarrow \mathring{\Delta}_c$ be a smooth map with θ -coordinate functions ϑ^i and let $\mathfrak{s} := \Lambda \circ \mathfrak{P} : X \rightarrow S^c$ the mapping of \mathfrak{P} to the 2-sphere. Then one has

$$\int_X \mathfrak{g}_{ij} \langle d\vartheta^i, d\vartheta^j \rangle_\delta d^2x = \int_X \|d\mathfrak{s}\|^2 d^2x \quad (4.101)$$

Proof. The sphere metric is induced by the Euclidean metric of the ambient \mathbb{R}^c . Accordingly, the Riemannian norm of a vector $v \in \mathbb{R}^c$ tangent to $2S^c$ is given by $\delta_{ij} v^i v^j$. For a vector $w \in \mathbb{R}^c$ tangent to $\mathring{\Delta}_c$, with coordinate representation $\mathbf{w} \in \mathbb{R}^c$, its Fisher-Rao norm at a point $p \in \mathring{\Delta}_c$ is given by

$$\frac{\delta_{ij} w^i w^j}{p_i} = \mathfrak{g}(p)_{ij} \mathbf{w}^i \mathbf{w}^j. \quad (4.102)$$

The isometry relation yields

$$\mathfrak{g}(p)_{ij} \mathbf{w}^i \mathbf{w}^j = \delta_{ij} v^i v^j \quad \text{for} \quad v^i = \frac{\partial \Lambda^i}{\partial p^j} w^j. \quad (4.103)$$

Specifically, if $\mathbf{w} = \partial_\mu \vartheta$, where ϑ is the coordinate vector associated to \mathfrak{p} , then $v = \partial_\mu \mathfrak{s}$ by the chain rule. Hence

$$\int_X \mathfrak{g}_{ij} \langle d\vartheta^i, d\vartheta^j \rangle_\delta d^2x = \int_X \mathfrak{g}_{ij} \partial_\mu \vartheta^i \partial_\nu \vartheta^j \delta^{\mu\nu} d^2x = \int_X \mathfrak{g}_{ij} \mathbf{w}^i \mathbf{w}^j \delta^{\mu\nu} d^2x \quad (4.104a)$$

$$\stackrel{(4.103)}{=} \int_X \delta_{ij} v^i v^j \delta^{\mu\nu} d^2x = \int_X \delta_{ij} \partial_\mu \mathfrak{s}^i \partial_\nu \mathfrak{s}^j \delta^{\mu\nu} d^2x \quad (4.104b)$$

$$= \int_X \|\mathrm{d}\mathfrak{s}\|^2 d^2x. \quad (4.104c)$$

□

Remark 4.23 (generalized S flows). All the steps above still work if we replace (X, δ) with another Riemannian manifold (M, h) . This paves the way for extensions of the S flow to more general base manifolds.

Corollary 4.24 (entropic harmonic energy on the sphere). Let $\mathfrak{P} : M \rightarrow \mathring{\Delta}_c$ be a smooth map and $\mathfrak{s} = \Lambda \circ \mathfrak{p} : M \rightarrow S^c$ the associated sphere-valued map. Then the entropic harmonic energy takes the form

$$E_\varphi(\mathfrak{P}) = \int_X \left(\frac{1}{2} \mathfrak{g}_{ij} \langle d\vartheta^i, d\vartheta^j \rangle_\delta - \varphi(\mathfrak{P}) \right) d^2x = \int_X \left(\frac{1}{2} \|\mathrm{d}\mathfrak{s}\|^2 - \varphi(\mathfrak{P}) \right) d^2x. \quad (4.105)$$

We are now in the position to compare the energy E_φ with the functional (2.31)

$$E_S(\mathfrak{P}) = \frac{1}{2} \int_X (\|\mathrm{d}\mathfrak{P}\|^2 - \|\mathfrak{P}\|^2) d^2x \quad (4.106)$$

governing the S flow (recall Section 2.3). The two functionals formulate similar goals in that, by minimization, both enforce smoothness of the function \mathfrak{p} while also penalizing configurations that are close to the barycenter in $\mathring{\Delta}_c$ with \mathfrak{s} the point $\theta = 0$. The S flow energy pushes to the boundary by maximizing the purity $\|p\|^2$, while the sigma flow energy maximize φ which means to minimize the entropy H . While the S flow energy enforces smoothness by reducing the L^2 -norm of the gradient of \mathfrak{P} , the sigma flow reduces the magnitude of the gradient of the corresponding sphere-valued map \mathfrak{s} .

Thus, both processes induced by minimizing the respective functionals achieve the same goal, namely to generate smooth mappings to the boundary of the simplex, yet in a slightly different manner.

4.6 Tangent Space Parametrization

In this section, we generalize the tangent space parametrization of assignment flows from [ZSPS20] to entropic sigma flows (σ - φ -flow) and an arbitrary α -connection, i.e. for the entropic extension of sigma- α flows (σ - α flow) (Def. 4.13). Such parametrizations are essential for numerical computation. We first introduce a convenient representation of the flow extending the tension field (4.10) and some further notation.

Definition 4.25 (entropic sigma- α flow). Let $\mathfrak{P} \in C^\infty(M, \mathring{\Delta}_c)$ and $\alpha \in \mathbb{R}, m^2 \in \mathbb{R}_>$ be given. The entropic sigma- α flow in θ -coordinates is the system

$$\mathfrak{S} \in C^\infty(M_T, \mathring{\Delta}_c) \quad \text{s.t.} \quad \begin{cases} \partial_t \vartheta^i &= \Delta_{h_t} \vartheta^i + \frac{1-\alpha}{2} \left(\langle d(\vartheta^i - \Psi), d(\vartheta^i - \Psi) \rangle_{h_t} \right. \\ &\quad \left. - \langle d\Psi, d\Psi \rangle_{h_t} \right) + m^2 \vartheta^i \\ h_t &= \mathcal{O}(\mathfrak{S}_t) \\ \mathfrak{S}_0 &= \mathfrak{P}, \end{cases} \quad (\sigma\text{-}\varphi\text{-}\alpha \text{ flow})$$

$\forall t \in (0, T), i \in [c]$, where \mathcal{O} satisfies the uniform positive definiteness condition (4.26) and ϑ^i are the coordinate functions of \mathfrak{S}_t and

$$\Psi := \psi \circ \mathfrak{S}_t : M \rightarrow \mathbb{R}. \quad (4.107)$$

For any vector in $\theta \in \mathbb{R}^c$, we denote by $(0, \theta) \in \mathbb{R}^c$ the vector with 0 prepended as first component. Furthermore, the natural logarithm \log as well as the exponential function apply componentwise to vectors, that is for a vectors $p \in \mathbb{R}_{>}^c$ and $v \in \mathbb{R}^c$, one has $(\log p)_i = \log p_i$ and $(e^v)_i = e^{v_i}$, for every $i \in [c]$. We set $\mathbf{1}_c := (1, \dots, 1)^\top \in \mathbb{R}^c$. The tangent space to $\mathring{\Delta}_c$ is the linear subspace

$$T_0 := \left\{ v \in \mathbb{R}^c : \sum_{i \in [c]} v_i = 0 \right\} \quad (4.108)$$

with linear orthogonal projection

$$\Pi_0 : \mathbb{R}^c \rightarrow T_0, \quad v \mapsto v - \left(\frac{1}{c} \sum_{i \in [c]} v_i \right) \mathbf{1}_c. \quad (4.109)$$

The mapping

$$\text{sm} : T_0 \rightarrow \mathring{\Delta}_c, \quad v \mapsto \text{softmax}(v) := \frac{e^v}{\langle \mathbf{1}_c, e^v \rangle} \quad (4.110)$$

defines a smooth diffeomorphism between T_0 and $\mathring{\Delta}_c$ with inverse given by

$$\text{sm}^{-1} : \mathring{\Delta}_c \rightarrow T_0, \quad p \mapsto \Pi_0 \log p. \quad (4.111)$$

Thus, we can uniquely parametrize elements p of $\mathring{\Delta}_c$ by tangent vectors v through the relation $p = \text{sm}(v)$. We consider furthermore the *replicator mapping* (as special case of the replicator tensor (2.25))

$$R : \mathring{\Delta}_c \rightarrow \mathbb{R}^{c \times c}, \quad p \mapsto R_p = [p_i(\delta_{ij} - p_j)]_{i,j \in [c]_0}. \quad (4.112)$$

It has the properties

$$R_p \Pi_0 = R_p, \quad \forall p \in \mathring{\Delta}_c \quad \text{and} \quad d \text{sm}(v) = R_{\text{sm}(v)}. \quad (4.113)$$

The maps R and Π_0 can be extended to a maps $R : C^\infty(M, \mathring{\Delta}_c) \rightarrow C^\infty(M, \mathbb{R}^{c \times c})$ and $\Pi_0 : C^\infty(M, \mathbb{R}^c) \rightarrow C^\infty(M, T_0)$ by post-composition.

We specify the relations between the tangent space parametrization and θ -coordinates.

Proposition 4.26 (tangent coordinates and θ -coordinates). For $p \in \mathring{\Delta}_c$ with coordinate representation (recall (2.11))

$$\theta = (\theta^1, \dots, \theta^c) = \chi_e(p) \quad (4.114)$$

and tangent space parameters

$$v = (v^0, \dots, v^c) = \text{sm}^{-1}(p) \in \mathbb{R}^c, \quad (4.115)$$

one has the relations

$$v^i = \theta^i + v^0, \quad i \in [c] \quad \text{where} \quad v^0 = -\frac{1}{c} \sum_{j \in [c]} \theta^j. \quad (4.116)$$

These relations are summarized by the equations

$$v = \Pi_0(0, \theta), \quad (4.117a)$$

$$(0, \theta) = v - v^0 \mathbf{1}_c. \quad (4.117b)$$

Proof. Equations (4.115), (4.111) and (4.109) yield

$$v^i = \log p_i - \frac{1}{c} \sum_{j \in [c]_0} \log p_j = \log \frac{p_i}{p_0} - \frac{1}{c} \sum_{j \in [c]_0} \log \frac{p_j}{p_0} \stackrel{(2.11)}{=} \theta^i - \frac{1}{c} \sum_{j \in [c]} \theta^j, \quad i \in [c]. \quad (4.118)$$

The expression for v^0 follows from $v \in T_0$ and (4.108):

$$v^0 = -\sum_{i \in [c]} v^i = \frac{c}{c} \sum_{j \in [c]} \theta^j - \sum_{i \in [c]} \theta^i = \frac{c-c}{c} \sum_{j \in [c]} \theta^j = -\frac{1}{c} \sum_{j \in [c]} \theta^j. \quad (4.119)$$

□

We extend the bilinear pairing on the cotangent space $\langle \cdot, \cdot \rangle_h : \Omega^1(M) \otimes \Omega^1(M) \rightarrow C^\infty(M)$ to a bilinear mapping on vector valued forms $\langle \cdot, \cdot \rangle_h : (\Omega^1(M) \otimes \mathbb{R}^c) \otimes (\Omega^1(M) \otimes \mathbb{R}^c) \rightarrow C^\infty(M, \mathbb{R}^c)$. For a smooth vector-valued function $f : M \rightarrow \mathbb{R}^c$ with $df \in \Omega^1(M) \otimes \mathbb{R}^c$, the pairing is locally given by

$$\langle \langle df, df \rangle_h \rangle(x) = (\partial_\mu f^i(x) \partial_\nu f^j(x) h^{\mu\nu}(x))_{i \in [c]_0} \quad x \in M. \quad (4.120)$$

With this notation, we describe the tangent space parametrization of the sigma flow model as follows.

Proposition 4.27 (tangent space representation of sigma flow). *Let $T > 0$ be given and $\mathfrak{S} : M_T \rightarrow \mathring{\Delta}_c$ be a solution to the entropic sigma- α flow system (σ - φ - α flow) for an initial condition \mathfrak{P} , the mass parameter $m^2 > 0$ and some fixed $\alpha \in \mathbb{R}$. Let $\mathbf{v}_t = \text{sm}^{-1}(\mathfrak{S})$ denote the tangent space representation of \mathfrak{S} . Then \mathbf{v} satisfies the PDE*

$$\frac{\partial}{\partial t} \mathbf{v}_t = \Pi_0 \left(\Delta_{h_t} \mathbf{v}_t + \frac{1-\alpha}{2} \langle d \log \text{sm}(\mathbf{v}_t), d \log \text{sm}(\mathbf{v}_t) \rangle_{h_t} + m^2 \mathbf{v}_t \right), \quad h_t = \mathcal{O}(\text{sm}(\mathbf{v}_t)), \quad (4.121)$$

where Π_0 acts on the function to its right by post-composition and the second term in the parenthesis is the application of the bilinear pairing $\langle \cdot, \cdot \rangle_h$ (4.120) to the vector valued 1-form $d \log \text{sm}(\mathbf{v}_t) \in \Omega^1(M) \otimes \mathbb{R}^c$. Furthermore, the function $\mathfrak{S} : M_T \rightarrow \mathring{\Delta}_c \subset \mathbb{R}^c$ satisfies the PDE

$$\frac{\partial}{\partial t} \mathfrak{S}_t = R_{\mathfrak{S}_t} \left(\Delta_{h_t} \log \mathfrak{S}_t + \frac{1-\alpha}{2} \langle d \log \mathfrak{S}_t, d \log \mathfrak{S}_t \rangle_{h_t} + m^2 \log \mathfrak{S}_t \right), \quad h_t = \mathcal{O}(\mathfrak{S}_t). \quad (4.122)$$

where the matrix function $R_{\mathfrak{S}_t}$ acts on the function to its right by pointwise matrix-vector multiplication and the second term in the parenthesis is the application of the bilinear pairing $\langle \cdot, \cdot \rangle_{h_t}$ (4.120) to the vector valued 1-form $d \log \mathfrak{S}_t \in \Omega^1(M) \otimes \mathbb{R}^c$.

Proof. In the following, we simplify notation and do not indicate the time dependencies $t \mapsto \mathbf{v}_t, t \mapsto h_t$ and $t \mapsto \vartheta_t$. From relation (4.117) follows

$$\frac{\partial}{\partial t} \mathbf{v} = \Pi_0 \frac{\partial}{\partial t} (0, \vartheta). \quad (4.123)$$

Then, due to (σ - φ - α flow)

$$\frac{\partial}{\partial t} \vartheta^i = \Delta_h \vartheta^i + \frac{1-\alpha}{2} (\langle d(\vartheta^i - \Psi), d(\vartheta^i - \Psi) \rangle_h - \langle d\Psi, d\Psi \rangle_h) + m^2 \vartheta^i \quad (4.124)$$

We denote the ambient coordinates of \mathfrak{S} by $(\mathfrak{p}_0, \dots, \mathfrak{p}_c)$. By virtue of (4.116), (2.12) and (2.14) we have

$$\mathfrak{S} = (\mathfrak{p}_0, \dots, \mathfrak{p}_c), \quad (\vartheta^i - \Psi) = \log \mathfrak{p}_i, \quad -\Psi = \log \mathfrak{p}_0, \quad \vartheta^i = \mathbf{v}^i - \mathbf{v}^0. \quad (4.125)$$

This yields

$$\frac{d}{dt} \vartheta^i = \Delta_h (\mathbf{v}^i - \mathbf{v}^0) + \frac{1-\alpha}{2} (\langle d \log \mathfrak{p}_i, d \log \mathfrak{p}_i \rangle_h - \langle d \log \mathfrak{p}_0, d \log \mathfrak{p}_0 \rangle_h) + m^2 (\mathbf{v}^i - \mathbf{v}^0) \quad (4.126)$$

and in turn

$$\frac{d}{dt} \vartheta^i = \Delta_h \mathbf{v}^i + \frac{1-\alpha}{2} \langle d \log \mathfrak{p}_i, d \log \mathfrak{p}_i \rangle_h + m^2 \mathbf{v}^i - \Delta_h \mathbf{v}^0 - \frac{1-\alpha}{2} \langle d \log \mathfrak{p}_0, d \log \mathfrak{p}_0 \rangle_h - m^2 \mathbf{v}^0. \quad (4.127)$$

The projection Π_0 is linear, fulfills $\Pi_0 \mathbf{1}_c = 0$ and allows to write

$$\Pi_0 \frac{\partial}{\partial t} (0, \vartheta) = \Pi_0 \left(\frac{\partial}{\partial t} (0, \vartheta) + [\Delta_h \mathbf{v}^0 + \frac{1-\alpha}{2} \langle d \log \mathfrak{p}_0, d \log \mathfrak{p}_0 \rangle_h + m^2 \mathbf{v}^0] \mathbf{1}_c \right), \quad (4.128)$$

which by (4.123) and (4.127) amounts to

$$\frac{\partial}{\partial t} \mathbf{v} = \Pi_0 \left(\Delta_h \mathbf{v} + \frac{1-\alpha}{2} \langle d \log \text{sm}(\mathbf{v}), d \log \text{sm}(\mathbf{v}) \rangle_h + m^2 \mathbf{v} \right). \quad (4.129)$$

Using (4.113) and the defining relation $\text{sm}(\mathbf{v}) = \mathfrak{S}$, the PDE governing \mathfrak{S} reads

$$\frac{\partial}{\partial t} \mathfrak{S} = R_{\mathfrak{S}} \dot{\mathbf{v}} = R_{\mathfrak{S}} \Pi_0 \left(\Delta_h \mathbf{v} + \frac{1-\alpha}{2} \langle d \log \text{sm}(\mathbf{v}), d \log \text{sm}(\mathbf{v}) \rangle_h + m^2 \mathbf{v} \right). \quad (4.130)$$

Using $R_{\mathfrak{S}} \Pi_0 = R_{\mathfrak{S}}$ (cf. (4.113)) and $\mathfrak{S} = \Pi_0 \log \mathbf{v}$ finally leads to

$$\frac{\partial}{\partial t} \mathfrak{S} = R_{\mathfrak{S}} \frac{\partial}{\partial t} \mathbf{v} = R_{\mathfrak{S}} \left(\Delta_h \log \mathfrak{S} + \frac{1-\alpha}{2} \langle d \log \mathfrak{S}, d \log \mathfrak{S} \rangle_h + m^2 \log \mathfrak{S} \right). \quad (4.131)$$

□

Remark 4.28 (advantage of tangent space parametrization). Regarding nonlinear flow integration on the probability simplex $(\hat{\Delta}_c, \mathfrak{g})$, the tangent parametrization enables better conditioned numerical computation than the exponential θ -coordinates.

5 Experiments and Comparison

This section presents few proof-of-concept experimental results that illustrate the sigma flow model.

- Section 5.1 describes a discretization of the approach (Definition 4.25) based on the tangent space parametrization (Proposition 4.27) and implementation details.
- Section 5.2 compares the resulting semi-discrete sigma flow model and the discretized S flow model from Section 2.3.
- The convergence behavior of the numerically integrated sigma flow with and without mass parameter is reported in Section 5.4, which illustrates Proposition 4.10 and Theorem 4.19.

Finally, we focus on the expressivity of the sigma flow model and on learning a mapping from data to the Riemannian metric h of the domain manifold M .

- Section 5.5 demonstrates that, for any given image, a metric exists which generates the image just out of noise. This raises the question to what extent such metric-valued mappings generalize to an entire *class* of images.
- We demonstrate empirically in Section 5.6 for computer generated labelings of *unseen noisy 2D random* polygonal regions, that this is indeed possible using a metric-valued mapping parametrized by a small neural net, even when then input is corrupted with a high level of noise. Details of the implementation are listed in Appendix A.2.

This finding sheds light on the intriguing problem of the generalization of this generative approach, which is just based on predicting a section of a 2×2 positive definite tensor field, to classes of *real* images. A corresponding thorough investigation is beyond the scope of this paper, however.

We therefore touch only briefly on this subject and conclude by illustrating how much the aforementioned map, trained on *synthetic* random Voronoi partitions, *fails* to generalize to *unconstrained real* images. Although the random polygonal scene scenario, that was used for learning, considerably differs from real images, it turns out that real image structure can be recovered remarkably well. Enhancing the metric prediction map which parametrizes the Laplace-Beltrami operator in order to close this gap, is left for future work.

5.1 Implementation

Numerical computations are based on the semi-discrete problem: the data manifold M is discretized but the time dimension is kept continuous. We assume that M is covered by a single global coordinate chart which is the case, e.g., for the torus $M = \mathbb{T}^2$.

The equation to be discretized is the sigma flow model with respect to α -connections and entropic potential with weight $m^2 \geq 0$ in ambient coordinates, i.e. the initial value PDE problem

$$\mathfrak{S} \in C^\infty(M_T, N), \quad \text{s.t.} \quad \begin{cases} \frac{\partial}{\partial t} \mathfrak{S} &= R_{\mathfrak{S}_t} \left(\Delta_{h_t} \log \mathfrak{S}_t + \frac{1-\alpha}{2} \langle d \log \mathfrak{S}, d \log \mathfrak{S} \rangle_h + m^2 \log \mathfrak{S} \right), \\ h_t &= \mathcal{O}(\mathfrak{S}_t), \\ \mathfrak{S}_0 &= \mathfrak{P}, \end{cases} \quad (5.1)$$

$\forall t \in (0, T)$, for a given $\mathfrak{P} \in C^\infty(M, \mathring{\Delta}_c)$ and some chosen fixed time period $[0, T]$, $T > 0$.

Spatial discretization entails to replace paths $[0, T] \rightarrow C^\infty(M, \mathring{\Delta}_c)$ by paths taking values in the assignment manifold, denoted by

$$\mathbb{S}: [0, T] \rightarrow \mathcal{W}_c^N, \quad (5.2)$$

where N is the number of points used to discretize M . Due to our assumption that M is covered by a single coordinate chart, we can discretize M by a regular grid. $\log \mathfrak{S}$ is discretized by $t \mapsto \log \mathbb{S}(t)$ where \log acts on all the entries of the matrix $\mathbb{S}(t)$ separately. The metric $h \in \Gamma(T^*(M_T) \otimes T^*(M_T))$ (recall the notation (3.11)) is represented by a matrix function

$$\mathbf{h}: [0, T] \rightarrow \mathbb{R}^{N \times 2 \times 2}. \quad (5.3)$$

We denote the first dimension slices of the tensor $\mathbf{h} \in \mathbb{R}^{N \times 2 \times 2}$ by

$$\mathbf{h}(t)_a \in \mathbb{R}^{2 \times 2}, \quad a \in [N]. \quad (5.4)$$

Discretization of the operator \mathcal{O} yields the function

$$\mathbb{O}: \mathcal{W}_c^N \rightarrow \mathbb{R}^{N \times 2 \times 2} \quad (5.5)$$

to be further specified below. The Laplace-Beltrami operator

$$\Delta_h = \frac{1}{\sqrt{|h|}} \partial_\mu \left(\sqrt{|h|} h^{\mu\nu} \partial_\nu \right) \quad (5.6)$$

is the composition of the differential operator $\partial_\mu (\sqrt{|h|} h^{\mu\nu} \partial_\nu)$ with the multiplication of the scalar function $1/\sqrt{|h|}$, which can be discretized individually. As for the differential operator, we adopt the discretization of [Wei98, Sec. 3.4.2] which yields a sparse matrix $\mathbb{E}_h \in \mathbb{R}^{N \times N}$. Discretizing the scalar function $1/\sqrt{|h|}: M \rightarrow \mathbb{R}$ yields a diagonal matrix $\mathbb{Q}_h \in \mathbb{R}^{N \times N}$ whose entries are given by

$$(\mathbb{Q}_h)_{aa} = \frac{1}{\sqrt{|h_a|}}. \quad (5.7)$$

As a result, the discretized Laplace-Beltrami operator is given by the matrix

$$\mathbb{L}_h = \mathbb{Q}_h \mathbb{E}_h \in \mathbb{R}^{N \times N}. \quad (5.8)$$

In view of the second term of the PDE (5.1), we describe now a discretization of the operation

$$\langle d \log \mathfrak{S}, d \log \mathfrak{S} \rangle_h = h^{\mu\nu} \partial_\mu (\log \mathfrak{S}) \partial_\nu (\log \mathfrak{S}) : M \rightarrow \mathbb{R}^c. \quad (5.9)$$

The derivative operators ∂_μ are discretized by stencil operators that generate matrices $\mathbb{D}_\mu \in \mathbb{R}^{N \times N}$, $\mu \in [\dim M]$. For instance, a 3×3 stencil for estimating the partial derivative ∂_1 in x -direction on a 2d grid is

$$\frac{1}{8} \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}. \quad (5.10)$$

The discrete derivative operators \mathbb{D}_μ act on the matrices $\log \mathbb{S} \in \mathbb{R}^{N \times c}$ by matrix multiplication from the left. As a result, the discretized version of (5.9) is given by the matrix

$$\langle \mathbb{D} \log \mathbb{S}, \mathbb{D} \log \mathbb{S} \rangle_h \in \mathbb{R}^{N \times c}, \quad \langle \mathbb{D} \log \mathbb{S}, \mathbb{D} \log \mathbb{S} \rangle_h)_a = h_a^{\mu\nu} (\mathbb{D}_\mu \log \mathbb{S})_a (\mathbb{D}_\nu \log \mathbb{S})_a, \quad a \in [N]. \quad (5.11)$$

We summarize these discretization rules in the following definition.

Definition 5.1 (semi-discrete sigma flow). Let $P \in \mathcal{W}_c^N$, $\alpha \in \mathbb{R}$, $m^2 \in \mathbb{R}_>$ be given. The *semi-discrete sigma flow model* is the initial value ODE problem

$$S : [0, T] \rightarrow \mathcal{W}_c^N \quad \text{s.t.} \quad \begin{cases} \dot{S} &= R_S \left(L_h \log S + \frac{1-\alpha}{2} \langle D \log S, D \log S \rangle_h + m^2 \log S \right), \\ h(t) &= O(S(t)), \quad \forall t \in [0, T], \\ S(0) &= P. \end{cases} \quad (5.12)$$

The first equation is understood as equality of functions $[0, T] \rightarrow T\mathcal{W}_c^N$, whereas $O : \mathcal{W}_c^N \rightarrow \mathbb{R}^{N \times 2 \times 2}$ maps to the open cone of symmetric positive-definite matrices and satisfies the uniform positive definiteness condition (4.26).

The dynamical system (5.12) evolves on the assignment manifold.

The papers [ZSPS20] and [BCA+24] have established numerical methods for the geometric numerical integration of such systems that can also be employed here.

5.2 Comparison to the Discrete S Flow Model

We compare the discrete S flow model from Section 2.3 and the semi-discrete sigma flow model (5.12) with $m^2 = 1, \alpha = 1$. This makes sense, because the S flow is intimately related to the e-geometry of the probability simplex which is realized by $\alpha = 1$. The discrete S flow ODE is given by

$$\dot{S} = R_S(L_\Omega S + S), \quad (5.13)$$

whereas the sigma flow ODE with $\alpha = 1, m^2 = 1$ reads

$$\dot{S} = R_S(L_h \log S + \log S), \quad (5.14)$$

which reveals the similarity of these two models. Both systems couple replicator equations as the original assignment flow model, yet with different fitness functions. To be specific, the S flow has its Laplacian operator parametrized by a weight matrix Ω , whereas the sigma flow is parametrized by a spatially discretized Riemannian metric h . Furthermore, the fitness function of the sigma flow acts purely on the logarithms of S , whereas the S flow acts on S directly.

5.3 Relation with the Transformer Network Architecture

The sigma model has been introduced from a *optimize-then-discretize* viewpoint: Minimizing various generalized harmonic energies (e.g., (3.6), (3.22), (4.7), (4.82)) generate variants of the sigma model approach. Subsequently, after discretization (Section 5.1), the differential equations are numerically integrated to determine the sigma flow.

A viable alternative is the antipodal *discretize-then-optimize* viewpoint that first focuses on a discretization of the harmonic energy functional. We only provide here a brief account of the ideas and refer to [GLM20, WY23] and [DHLM05].

First, assume that the Riemannian manifold (M, h) has been approximated by a *simplicial complex*, which is plausible at least in the case of image data, see [GY02] for instance. This discretization may be represented in terms of a graph $G = (V, E)$ with $N = |V|$ vertices and a weight matrix $\Omega \in \mathbb{R}^{N \times N}$ satisfying

- non-negativity: $\Omega_{ab} \geq 0$, $a, b \in V$,
- symmetry: $\Omega_{ab} = \Omega_{ba}$, $\forall a, b \in V$,
- support on edges: $\Omega_{ab} > 0 \Leftrightarrow ab \in E$.

Section 5.1 provides a basic example: A regular grid is used to discretize a torus \mathbb{T}^2 , and a local 3×3 symmetric stencil is used to discretize a differential operator of the form $u \mapsto \text{div}(D\nabla u)$. This realizes an elementary simplicial complex where the graph is given by the grid and the weight matrix is encoded by the stencils weights. These weights matrix fulfills the three criteria for a weight matrix of [Wei98, pp. 76].

In this set-up, maps $\mathfrak{P} : M \rightarrow \mathring{\Delta}_c$ are discretized by assignment matrices $P \in \mathcal{W}_c^N$, and the harmonic energy functional is succinctly discretized to

$$E : \mathcal{W}_c^N \rightarrow \mathbb{R}, \quad P \mapsto \sum_{a,b \in [N]^2} \Omega_{ab} d_g(P_a, P_b), \quad (5.15)$$

where d_g denotes the distance function on $\mathring{\Delta}_c$ induced by the Fisher-Rao metric. The variation of this discrete harmonic energy is given by the *discrete tension field* [GLM20]

$$\Upsilon(\mathbf{s}, \Omega) \in \mathbb{R}^{N \times c}, \quad \Upsilon(\mathbf{s}, \Omega)_a = \frac{1}{\mu_a} \sum_{b \in [N]} \Omega_{ab} \exp_{P_a}^{-1}(P_b) \in T_0, \quad (5.16)$$

where μ_a is the vertex weight induced by the simplicial complex; see [DHLM05] for details. This expression can be simplified when we isometrically identify the simplex $\mathring{\Delta}_c$ with the intersection of the positive orthant and the sphere S^{c-1} ([ÅPSS17, Lemma 1]) and replace P by a matrix $\mathbf{s} \in \mathbb{R}^{N \times c}$ whose columns are unit vectors. The discrete tension field takes the explicit form

$$\Upsilon(\mathbf{s}, \Omega)_a = \frac{1}{\mu_a} \sum_{b \in [N]} \Omega_{ab} \frac{\arccos(\langle \mathbf{s}_a, \mathbf{s}_b \rangle)}{1 - \langle \mathbf{s}_a, \mathbf{s}_b \rangle} (\mathbf{s}_b - \mathbf{s}_a \langle \mathbf{s}_a, \mathbf{s}_b \rangle), \quad (5.17)$$

Using this representation, we can formulate an alternative discrete sigma flow equation with fixed weight matrix of the form

$$\dot{\mathbf{s}}(t) = \Upsilon(\mathbf{s}(t), \Omega), \quad \text{where } \mathbf{s}_a(t) \in S^{c-1} \text{ and } \mathbf{s}_a(t) \geq 0, \quad \forall a \in [N], t > 0. \quad (5.18)$$

To generalize, we can make the weight matrix Ω dependent on the state $\mathbf{s}(t)$ and the time t , i.e. we consider time-variant weight matrices. As a result, we obtain a dynamical system evolving on a product of unit spheres, whose solution can be approximated by the geometric Euler scheme (assuming no constraints are violated)

$$\mathbf{s}(t_{k+1}) = \Pi\left(\mathbf{s}(t_k) + \delta \Upsilon(\mathbf{s}(t_k), \Omega(\mathbf{s}(t_k), t_k))\right), \quad (5.19)$$

where $\Pi : \mathbb{R}^{N \times c} \rightarrow (S^{c-1})^N$, $x \mapsto x/\|x\|$ is the standard projection and δ is the step size for numerical integration. We analyze this update step in the extreme case $\langle \mathbf{s}_a, \mathbf{s}_b \rangle = 0$ corresponding to the largest possible distance on the part of the sphere contained in the positive orthant. The equations then simplify to

$$\mathbf{s}(t_{k+1}) = \Pi\left(\mathbf{s}(t_k) + \frac{\delta \pi}{2\mu} \Omega(\mathbf{s}(t_k), t_k) \mathbf{s}(t_k)\right) \quad (\text{equality of matrices}) \quad (5.20)$$

This expression shares many similarities with an *attention block* of the *Transformer* network architecture [VSP⁺17]. As presented in [GLPR23, pp. 5], a simplified version of the attention block can be abstractly understood as the update rule

$$\mathbf{s}(t_{k+1}) = \Pi\left(\mathbf{s}(t_k) + \text{SelfAttention}(\mathbf{s}(t_k)) \mathbf{s}(t_k)\right), \quad (5.21)$$

where the so-called *dot product self-attention mechanism*

$$\text{SelfAttention}(\mathbf{s}(t_k)) = \text{sm}\left(\langle \mathbf{s}(t_k), \mathbf{s}(t_k) \rangle\right) \quad (5.22)$$

induces a valid symmetric weight matrix on a graph with N vertices that is fully connected. This simplified version of the attention block ignores fully connected layers usually employed in the transformer architecture. The comparison shows however that this version of the discrete sigma flow can be interpreted as a generalized attention block, since the formulation (5.18) realizes the simplified attention block as a special case on a fully connected graph. Furthermore, the choice

$$\Omega(\mathbf{s}(t_k), t_k) = \text{EdgeMask}(\text{SelfAttention}(\mathbf{s}(t_k))) \quad (5.23)$$

would be a valid weight matrix if the entries of the self-attention block not attending to connected vertices are masked out. On the other hand, the formulation (5.18) contains only weights on graph edges, because the discretization scheme is only supposed to achieve the first order approximation of the smooth structure. In order to approximate the smooth structures to higher precision, accounting for the contribution of points in two-hop neighborhoods etc. would be required. This produces a fully supported weight matrix in the limit of highest order approximation with a fixed number of discretization points.

We leave the investigation of mutual connections and implications for research from the continuous *optimize-then-discretize* perspective, as adopted in this paper, for future work.

5.4 Synthetic Benchmark: Convergence Behavior

We demonstrate the convergence behavior of the sigma flow model using the following set-up. The torus $M = \mathbb{T}^2$ serves as base manifold, with the standard flat metric $h = \delta$ and the simplex $\mathring{\Delta}_4$ as the target manifold. We used an initial configuration \mathfrak{P} given by

$$\mathfrak{P} : \mathbb{T}^2 \rightarrow \mathring{\Delta}_4, \quad (\xi_1, \xi_2) \mapsto \text{sm} \begin{pmatrix} x = 0.2(3 + \cos \xi_1) \cos \xi_2 \\ y = 0.2(3 + \cos \xi_1) \sin \xi_2 \\ z = 0.2 \sin \xi_1 \\ x + y + z \end{pmatrix}, \quad (5.24)$$

where sm denotes the softmax function $\mathbb{R}^4 \rightarrow \mathring{\Delta}_4$ defined by (4.110). Figure 5.1a shows a plot of $\chi_m \circ \mathfrak{P}$.

We first examined with $m^2 = 0$ the solution to the PDE system as special case of the sigma model

$$\frac{\partial}{\partial t} \mathfrak{S}_t = \tau(\mathfrak{S}_t, \mathfrak{g}, \delta), \quad \mathfrak{S}_0 = \mathfrak{P}, \quad (5.25)$$

which was discretized and numerically solved as explained in Section 5.1. The torus $\mathbb{T}^2 = S^1 \times S^1$ was discretized by a regular grid with periodic boundary conditions.

Figure 5.1b shows \mathfrak{S}_t for different times t . \mathfrak{S}_t converges to a point for large integration times which corroborates the statement of Proposition 4.10. We point out that we never had to resort to the regularized metric (4.21), that is the flow computation converged for $\varepsilon = 0$. We repeated the experiment with $m^2 = 1$. The trajectories emanating from the initial toroidal surface are shown by Figure 5.1c. All integral lines of the flow lie approximately on the *Wright manifold* [HS98, Sec. 18.8], a 2 dimensional submanifold of $\mathring{\Delta}_4$. This illustrates that solutions to the sigma flow equations are capable of preserving non-trivial relations during time evolution. Although no metric regularization ε was used, convergence to the boundary of the simplex was always observed. Varying α did not change the convergence behavior, but the speed of convergence as shown by Figures 5.1e and 5.1f, respectively.

5.5 Expressivity of the Sigma Flow Model

We studied the capacity of the entropic sigma flow (σ - φ -flow)

$$\frac{\partial}{\partial t} \mathfrak{S} = \tau_{\varphi}(\mathfrak{S}, \mathfrak{g}, h), \quad \mathfrak{S}_0 = \mathfrak{P} \quad (5.26)$$

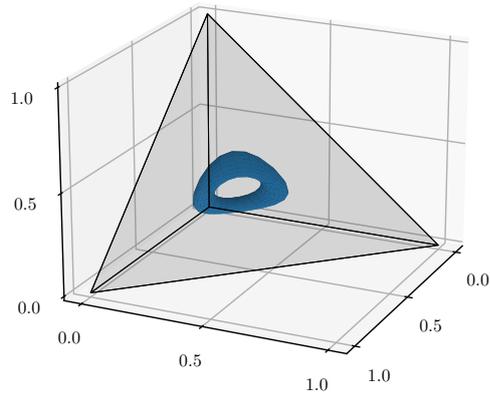
to parametrize a labeling

$$\mathfrak{L} : \mathbb{T}^2 \rightarrow [20] \quad (5.27)$$

using a *fixed* metric h on $M = \mathbb{T}^2$, as a basis for predicting such mappings in terms of a corresponding metric using given data (Section 5.6).

To this end, we solved the optimization problem

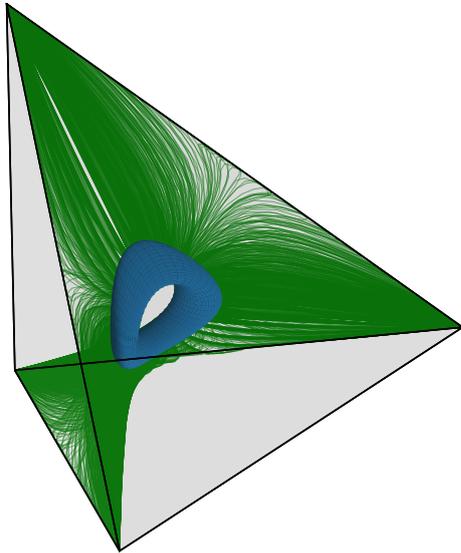
$$\min_h \int_{\mathbb{T}^2} \text{KL}[\mathfrak{S}_T(x) : \mathfrak{L}(x)] d^2x \quad \text{s.t.} \quad \mathfrak{S} \in C^\infty(M_T, \mathring{\Delta}_c), \quad \frac{\partial}{\partial t} \mathfrak{S}_t = \tau_{\varphi}(\mathfrak{S}_t, \mathfrak{g}, h), \quad \mathfrak{S}_0 = \mathfrak{P}, \quad (5.28)$$



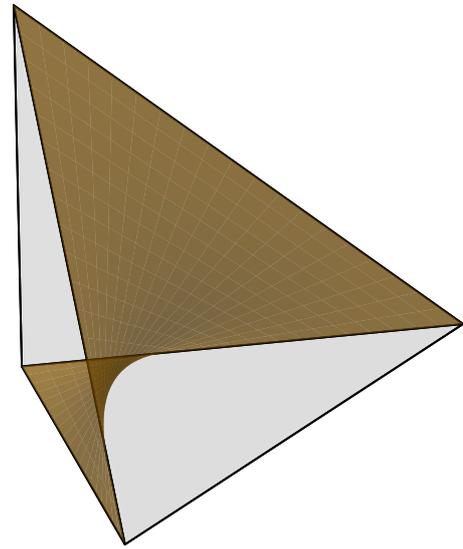
(a) Visualization of the torus inside $\hat{\Delta}_4$ using the chart χ_m .



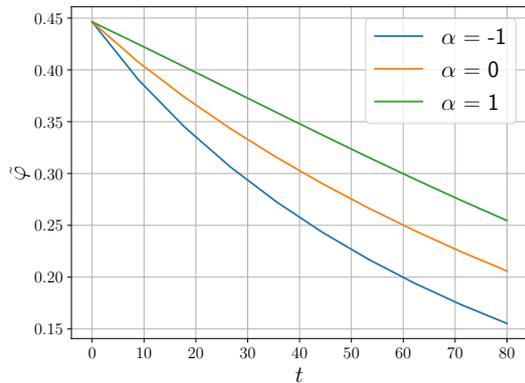
(b) Convergence of embedded torus to a point ($m = 0$).



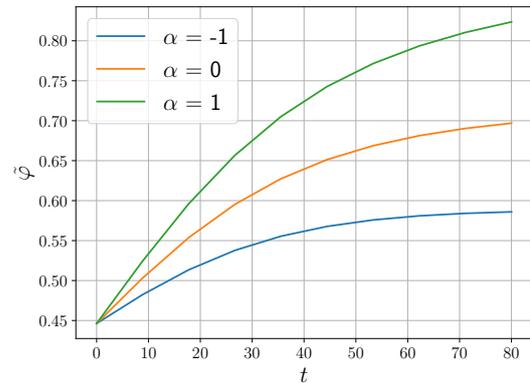
(c) Flow converges to simplex boundary almost everywhere when $m = 1$. Green lines represent the integral curves of the flow emanating from the initial torus.



(d) The entropic sigma flow approximately preserves the *Wright manifold* [HS98, Sec.18.8], a non-trivial submanifold of the simplex.



(e) Speed of convergence to the center for different values of α and $m = 0$. $\tilde{\varphi}$ is φ normalized to the range $[0, 1]$.



(f) Speed of convergence to the boundary for different values of α and $m = 0.1$. $\tilde{\varphi}$ is φ normalized to the range $[0, 1]$.

Figure 5.1: Illustration of the convergence behavior of the sigma flow for a low dimensional toy example.

for a fixed initial configuration \mathfrak{P} (see Figure 5.2a). The optimization was performed using gradient descent and backpropagation. We repeated the optimization for two different target configurations \mathfrak{L} , displayed as Figures 5.2b and 5.2c, respectively¹. The label configurations are visualized with a color code.

Figure 5.2d shows the resulting state \mathfrak{S}_T at different stages of the optimization, demonstrating that the desired configuration has been gradually achieved. Optimization was carried out until convergence, which took 1000 steps for the simple configuration and 2000 steps for the complex configuration. Figures 5.2e and 5.2h show the results for \mathfrak{S}_T , with incorrectly labelled pixels marked with black. We observed that the approximation locally degrades at pixels within either high frequency textured regions or along sharp edges.

In order to facilitate the interpretation of the learned metrics h^{-1} , we decomposed the metric into two parts

$$h^{-1} = \frac{h^{-1}\sqrt{|h|}}{\sqrt{|h|}} \quad (5.29)$$

The term $h^{-1}\sqrt{|h|}$ yields the part which is conformally invariant (see (3.30)) whereas $1/\sqrt{|h|}$ represents a scale factor for the diffusion process. We visualize the final metric h^{-1} by plotting separately the Riemannian anisotropy index² [MB06, Sec. 17.3] of $h^{-1}\sqrt{|h|}$ and the scale factor $1/\sqrt{|h|}$ in Figures 5.2g, 5.2h and 5.2i, 5.2j for the two versions of the experiment respectively.

The metric is strongly anisotropic along the edges and more isotropic in regions with constant labeling. The scale shows the opposite behavior: it becomes small along the edges and large in constant regions (note the *inverted* color scale used for displaying anisotropy and scale, respectively). This behavior is natural: the metric guides the diffusion process such that information propagation across edges is suppressed, but reinforced in regions with constant labeling. If the local spatial frequency of texture is too large, however, the image structure cannot be modeled by diffusion anymore.

5.6 Learning the Prediction of Labelings in Terms of a Time-Variant Metric

The experiments reported in this section concern the *learned* operator

$$\mathcal{O} : C^\infty(\mathbb{T}^2 \times [0, T], \mathring{\Delta}_{20}) \rightarrow C^\infty(\mathbb{T}^2, \mathbb{R}^{2 \times 2}) \quad (5.30)$$

which parametrizes all variants of the sigma model in terms of the metric $h_t = \mathcal{O}(\mathfrak{S}_t, t)$. Here the dependency of the operator \mathcal{O} on both the state $\mathfrak{P} : \mathbb{T}^2 \rightarrow \mathring{\Delta}_{20}$ and the time $t \in [0, T]$ was taken into account.

We selected a set of training data

$$\text{TS} = \{\mathfrak{L} : \mathbb{T}^2 \rightarrow [20]\} \quad (5.31)$$

consisting of pointwise labeled (segmented) tori with a random polygonal spatial structure. A color-coded sample from the training set is depicted by Figure 5.3a. The objective of training was to learn a *prediction map* (5.30) for the recovery of the ground truth labeling from a strongly corrupted input signal. We denote by Corr the distribution of corruptions applied to the data. Figure 5.3b shows an example of a corrupted sample from the training set.

We used a small neural network to define and parametrize a class of operators

$$\mathcal{C} \subset C^\infty(\mathbb{T}^2 \times [0, T], \mathring{\Delta}_{20}) \rightarrow C^\infty(\mathbb{T}^2, \mathbb{R}^{2 \times 2}). \quad (5.32)$$

The learning task was formulated as the optimization problem

$$\min_{\mathcal{O} \in \mathcal{C}} \sum_{\mathfrak{L} \in \text{TS}} \mathbb{E}_{\sigma \sim \text{Corr}} \left[\int_{\mathbb{T}^2} \text{KL}[\mathfrak{S}_T(x) : \mathfrak{L}(x)] d^2x \right] \quad \text{s.t.} \quad \frac{\partial}{\partial t} \mathfrak{S}_t = \tau_{\bar{\varphi}}(\mathfrak{S}_t, \mathfrak{g}, \mathcal{O}(\mathfrak{S}_t, t)), \quad \mathfrak{S}_0 = \sigma(\mathfrak{L}). \quad (5.33)$$

¹The second label configuration has been generated by clustering the RGB values of a realistic image of a mandrill by the k -means clustering algorithm with 20 cluster centers. The resulting labels represent a valid configuration $\mathfrak{L} : \mathbb{T}^2 \rightarrow [20]$ depicted in 5.2c.

²The *anisotropy index* measures the distance of a matrix to the closed isotropic positive definite matrix, inside the cone of symmetric positive definite matrices using the corresponding geodesic distance.



(a) Initial configuration \mathfrak{P} (left) with close-up view of the patch marked with red (right).

(b) Fixed label configuration with simple spatial structure.

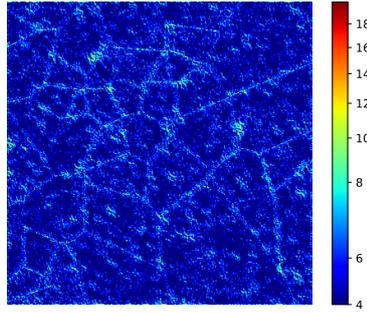
(c) Fixed label configuration with complex spatial structure.



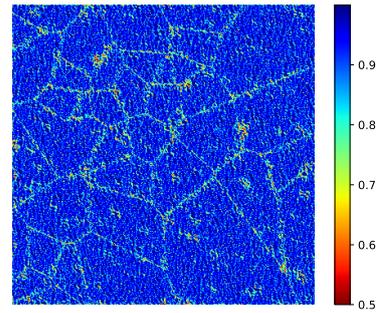
(d) Evolution of $\mathfrak{S}_{\mathcal{T}}$ after subsequent 100 gradient descent steps. The target labeling gradually emerges.



(e) The final state $\mathfrak{S}_{\mathcal{T}}$ corresponding to the evolution (d). Pixels with wrong label are marked black. The approximation is nearly exact.



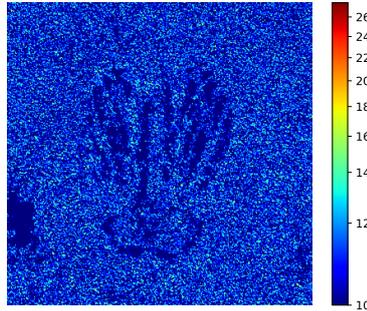
(f) Visualization of the anisotropy index of the conformally invariant part $h^{-1}\sqrt{|h|}$ of the fitted metric, after a histogram transformation to enhance visibility of the top range of values. Diffusion is strongly anisotropic along edges and at junctions of the target label configuration.



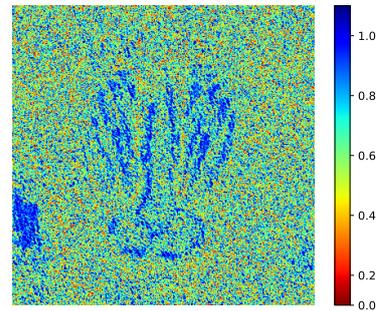
(g) Visualization of the scale factor $1/\sqrt{|h|}$ of the fitted metric, using the *inverted* color scale relative to (f). The scale decreases at edges suppresses diffusion across edges. Inside regions with constant labeling, the scale increases and diffusion progresses quickly.



(h) The final state $\mathfrak{S}_{\mathcal{T}}$ of the evolution analogous to (d). Pixels with wrong label are marked black. The approximation fails in non-smooth textured regions.



(i) The anisotropy index of $h^{-1}\sqrt{|h|}$ for the fitted metric h^{-1} , after a histogram transformation to enhance visibility of the top range of values. The same conclusion as for (f) can be drawn, where in this case the edge structure is highly irregular, however.



(j) The scale factor $1/\sqrt{|h|}$ of the fitted metric (inverted color scale). Similar to the anisotropy index of the metric (panel (i)), the scale increases (decreases) in regions with smooth (non-smooth) labeling.

Figure 5.2: Expressivity of sigma flow model depending on the fitted metric h .

In other words, the learned operator \mathcal{O} is supposed to accomplish the following. If \mathfrak{S}_0 is a corrupted initial configuration, then \mathfrak{S}_T should recover the uncorrupted configuration. In particular, this should hold for *unseen* and *independently sampled random* test data from the same image class. The optimization problem was solved through backpropagation and gradient descent.

Results of the test phase are shown by Figure 5.3c. In order to assess the effect of learning, we compared the outcome of diffusion with the predicted metric with the outcome of diffusion using the flat metric δ . We observed that the learned metric significantly improves the denoising capacities in comparison to the flat metric.

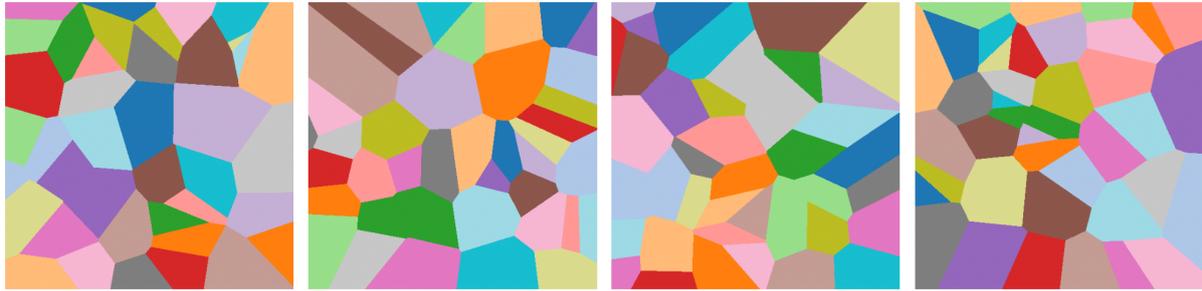
Furthermore, we tested the learned mapping (5.30) also on labeled real images \mathfrak{L} whose spatial structure differs considerably from those of the training images (Fig. 5.3d)³. Although the mapping \mathcal{O} is merely parametrized by a 2×2 metric tensor field, and by a corresponding small neural network that was separately applied at each pixel for metric prediction, generalization to such more complex image structure happened to some extent, in particular in regions with a spatial structure the resembles the training data.

In order to further interpret the output of the learned operator, we applied \mathcal{O} to an unperturbed label configuration \mathfrak{L} and analyzed the time-averaged metric

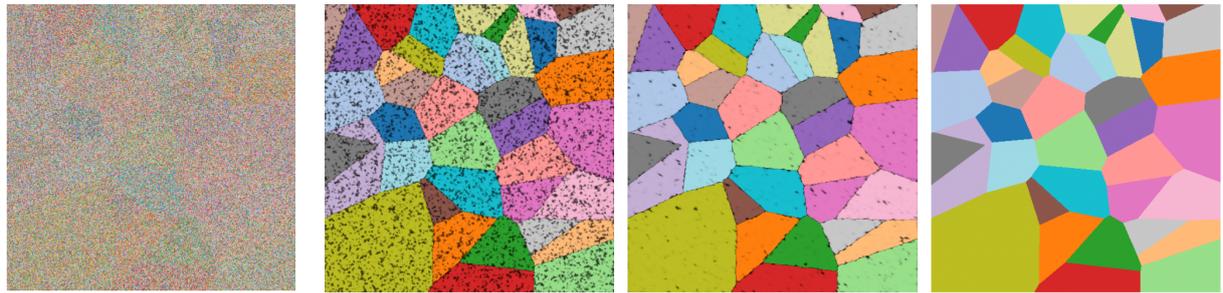
$$h^{-1} = \int_0^T \mathcal{O}(\mathfrak{L}, t) dt. \quad (5.34)$$

Analogous to Figure 5.2, Figure 5.3e shows the anisotropy index the unit determinant part $h^{-1} \sqrt{|h|}$ of the predicted metric, and Figure 5.3f shows the corresponding scale factor $1/\sqrt{|h|}$. The results show that the predicted metric h^{-1} generalizes fairly well the behavior shown by Figure 5.2 to such unseen novel test data.

³The labeling \mathfrak{L} for this test was generated by applying k -means clustering with 20 cluster centers to the RGB values of the image kodim20 from the Kodak database <https://r0k.us/graphics/kodak/>.



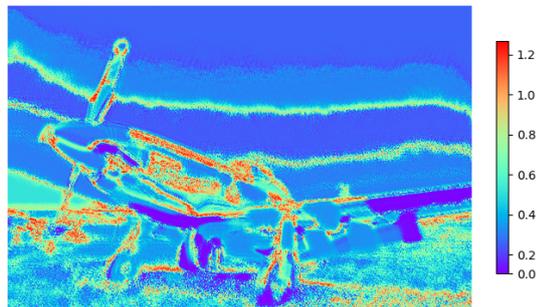
(a) Few labeling samples \mathcal{L} from the training set.



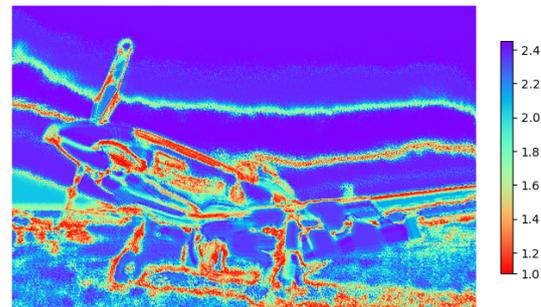
(b) Example of a corrupted independent test sample. The learning objective is to predict the labeling shown as right-most panel (c).
 (c) Test results of the learned operator \mathcal{O} for the input (b). Incorrectly labeled pixels are marked black. *Left:* Prediction $\mathfrak{S}_{\mathcal{T}}$ of the sigma model using the flat metric δ . *Center:* Prediction $\mathfrak{S}_{\mathcal{T}}$ using the learned metric \mathcal{O} improves the performance. *Right:* Ground truth for this test sample.



(d) Evaluation of the learned operator \mathcal{O} for a labeling dissimilar to the training set. Incorrectly labeled pixels are marked with black. *Left:* Prediction $\mathfrak{S}_{\mathcal{T}}$ of the sigma model using the flat metric δ . *Center:* Prediction $\mathfrak{S}_{\mathcal{T}}$ using the learned metric \mathcal{O} improves performance. Errors mostly occur at edges and in non-smooth regions. *Right:* Ground truth for this test sample.



(e) Anisotropy index of the conformally invariant part $h^{-1}\sqrt{|h|}$ of the time-averaged metric (5.34), evaluated using an *uncorrupted* test labeling \mathcal{L} . The absence of noise (in contrast to the training phase) slightly deteriorates prediction performance. Overall, however, the metric adapts well to oriented image structure.



(f) Scale factor $1/\sqrt{|h|}$ of the time-averaged metric (5.34), evaluated using an uncorrupted test labeling \mathcal{L} (inverted color scale). The scale factor consistently decreases close to edges and suppresses diffusion. Note the inverted color map with respect to 5.3e.

Figure 5.3: Testing the operator \mathcal{O} learned from simple label configurations using more complex configurations.

6 Conclusion

6.1 Summary

Sigma flow model. This paper introduced the sigma flow model for image and metric data labeling on graphs. The model is based on a generalized harmonic energy as objective function between a Riemannian domain and target manifold, respectively. Geometric integration of the Riemannian gradient flow optimizes the mapping. The flow is called ‘sigma flow’ because the approach resembles the mathematical structure of sigma models known in quantum field and string theory.

Specific choices of the domain and target manifold and the generalized harmonic energy yield variants of the sigma flow approach. We specifically focused on image data on a two-dimensional domain manifold and the probability simplex equipped with the Riemannian Fisher-Rao metric. This variant of the sigma flow model combines the Laplace-Beltrami approach to image denoising and enhancing introduced by Sochen, Kimmel and Malladi about 25 years ago [SKM98] and the assignment flow for metric data labeling developed by the authors. We proved that this sigma flow model constitutes a proper nonlinear geometric diffusion approach such that its variant based on the generalized entropic harmonic energy constitutes a proper labeling approach.

Two-stage parametrization of structured prediction via sigma flows. A remarkable feature of our geometric approach is the chain of self-referring time-variant parametrizations of large-scale structured prediction in terms of sigma flows, as sketched by Figure 1.1 and repeated here for the reader’s convenience:

$$\dots \mathfrak{S}_t \xrightarrow{\mathcal{O}} h_t \xrightarrow{\Delta_{h_t}} \tau(\mathfrak{S}_t, g, h_t) \xrightarrow{\int dt} \mathfrak{S}_t \dots \quad (6.1)$$

The tension field τ which governs the evolution of the state $t \mapsto \mathfrak{S}_t$ is parametrized by the Laplace-Beltrami operator Δ_{h_t} that itself is parametrized by the Riemannian metric h_t of the domain manifold. We showed that by making this metric dependent in terms of a mapping \mathcal{O} on both given data and the evolving state, our sigma flow model covers a range of established nonlinear PDE models of mathematical image analysis. In particular, the mapping \mathcal{O} can be parametrized by a neural network whose parameters can be conveniently learned from data, due to the inherent smoothness of our geometric approach and the robust numerics used for geometric integration of the sigma flow.

Expressivity of sigma flows and learning the generator from data. We demonstrated the remarkable expressivity of the sigma flow: any image structure can be generated from pure white noise by choosing properly the domain metric h , i.e. the parametrization of the generator Δ_h (Laplace Beltrami operator) of the generator $\tau(\Delta_h)$ (tension field) of the sigma flow. This suggests to study the sigma flow model from the viewpoint of machine learning, since the aforementioned succinct mathematical representation of structured prediction by nonlinear sigma flows should enable strong task-specific adaptivity by using a compact set of parameters learned from data.

We briefly demonstrated this property using a fairly small neural network for the parametrization of the mapping \mathcal{O} , that generates the domain metric h_t as ‘seed’ of the sigma flow, according to (6.1). As proof of concept, we showed empirically that learning in this way the generation of a field of 2×2 metric tensors (i.e. the evolving discretized metric h_t) enables to cope with a class of random polygonal scenes, even when contaminated with a high level of noise. Furthermore, applying this trained model directly to more general labelings of real images deteriorates prediction performance, but does not cause it to break down.

6.2 Further Work

The last statement raises the question: how rich have parametrizations of the time-variant metric h_t to be in order to generalize the labeling prediction performance to classes of real images? We point out that this specific question differs from the general problem of the role of the parameters of general black-box deep networks, since

- (i) a major part of the prediction performance is already ‘hard-wired’ by the geometric structure of sigma flows and by geometric integration, which generates the ‘layers’ of a corresponding deep network;

(ii) the parametrization only concerns the ‘generator of the generator’ of sigma flows, as discussed above in connection with (6.1).

These issues regarding the feasibility of low-dimensional parametrizations should also stimulate probabilistic extensions towards *generative models* for *discrete* structured predictions, along the lines recently worked out using assignment flows [BGAPS24].

Another aspect of our work concerns the appearance of ingredients of the transformer network architecture, as outlined in Section 5.3, by adopting a ‘discretize first – then optimize’ approach. Since we adopted the antipodal ‘optimize first – then discretize’ approach which better conforms to the established mathematical structures of PDE-based data analysis and mathematical physics, we hope that our work contributes to the field of geometric deep learning and to promoting geometric design principles in other areas of scientific machine learning.

A Appendix

A.1 Glossary of Symbols

Normal Font symbols

M	smooth closed oriented Riemannian manifold, image domain
N	smooth manifold <i>and</i> number of vertices of a grid graph
h	Riemannian metric on M
$ h $	determinant of h
g	Riemannian metric on N
Γ	global sections of vector bundles on M
δ	Euclidean metric (3.14) <i>and</i> Kronecker delta
f	smooth function $M \rightarrow \mathbb{R}$
M_T	time cylinder $(0, T) \times M$ of M of length T (3.11)
ω_h	volume form of h on M (2.1a)
d	(de-Rahm) differential on M (3.2)
Ω	weight matrix for S flow (2.20) <i>and</i> domain in \mathbb{R}^2 for image functions (3.18)
X	open domain in \mathbb{R}^2 (2.31)
u	time dependent gray value image $u : \Omega_T \rightarrow \mathbb{R}$ (3.12)
f_t	for $f : M_T \rightarrow \mathbb{R}$ and $t \in (0, T)$: $f_t = f(t, \cdot)$ is the constant time slice of f
\mathbb{T}^2	two-dimensional torus $S^1 \times S^1$
$\tau(f, g, h)$	tension field (3.7)
E_H	harmonic energy (3.5)
Δ_h	Laplace-Beltrami operator (3.9)
ϕ_n	eigenfunctions of Δ_h (4.28)
λ_n	eigenvalues of Δ_h (4.28)
a_n	expansion coefficients of ϑ in the basis ϕ_n (4.64)
κ	warp factor for anisotropic diffusion (3.20)
B	scalar function for the generalized diffusion ansatz (Perona-Malik, TV) (3.22)
λ	conformal factor (3.30)
$C(\mathcal{O})$	uniform boundedness of \mathcal{O}
V	diffusion tensor (in the anisotropic diffusion setting) (3.18)
c_1, c_2	bounds for the matrix function \mathcal{B} (4.30)
m^2	mass parameter (4.16)
c	number of labels (categories, classes)
Δ_c	probability simplex in \mathbb{R}^c (2.9)
$\overset{\circ}{\Delta}_c$	relative interior of the simplex Δ_c (2.10)
χ_e, χ_m	coordinate charts on $\overset{\circ}{\Delta}_c$ (2.11)
p_i	m -coordinates on $\overset{\circ}{\Delta}_c$ (2.12)
θ_i	e -coordinates on $\overset{\circ}{\Delta}_c$ (2.12)
\mathfrak{c}	dimension $\mathfrak{c} = c - 1$ of $\overset{\circ}{\Delta}_c$; index range of local m - and e -coordinates (2.11)
T_0	tangent space to $\overset{\circ}{\Delta}_c$ (4.108)
v	tangent vector $v \in T_0$
Π_0	orthogonal tangent space projection (4.109)
R	replicator operator (2.25)
sm	softmax function (4.110)
α	parameter for α -connections on $\overset{\circ}{\Delta}_c$, see (2.19)
Γ	Christoffel symbols on $\overset{\circ}{\Delta}_c$ (2.18) <i>and</i> sections of vector bundles
φ	negative entropy (2.16)
H	entropy $H = -\varphi$ (2.30)
$\tilde{\varphi}$	shifted negative entropy (2.16)
ψ	convex (Legendre-Fenchel) conjugate $\psi = \varphi^*$ of φ (2.14)

E_S	continuous domain S flow energy (2.31)
$E_{\tilde{\varphi}}$	entropic harmonic energy (4.16)
Λ	sphere map (4.100)
KL	Kullback-Leibler divergence for probability vectors (2.16) and Kullback-Leibler divergence extended to assignment matrices (5.28)
TS	test set used in (5.31)
Corr	probability distribution used for corrupting labelings; cf. (5.33)

Symbols for smooth quantities

g	Fisher Rao metric (2.17c)
ε	parameter regularizing the Fisher-Rao-metric (4.21)
\mathfrak{d}	partial derivative along θ -coordinates on $\mathring{\Delta}_c$ (2.6)
\mathfrak{S}	smooth function $M_T \rightarrow \mathring{\Delta}_c$, smooth analog of assignment matrix S (4.4)
\mathfrak{P}	smooth function $\mathfrak{P}: M \rightarrow \mathring{\Delta}_c$, starting point of sigma flow, smooth analog of P (4.4)
\mathfrak{p}	m -coordinate expression $\mathfrak{p} = \chi_m \circ \mathfrak{S}_t$ of \mathfrak{S}_t (4.5)
\mathfrak{e}	e -coordinate expression $\mathfrak{e} = \chi_e \circ \mathfrak{S}_t$ of \mathfrak{S}_t (4.5)
\mathfrak{s}	\mathfrak{S}_t mapped $\mathfrak{s} = \Lambda \circ \mathfrak{S}_t$ to the sphere (4.101)
\mathfrak{v}	tangent space representation $\mathfrak{v} = \text{sm}^{-1} \circ \mathfrak{S}_t$ of \mathfrak{S}_t (Prop. 4.27)
Ψ	composition $\Psi = \psi \circ \mathfrak{S}_t$ of the Hessian potential of (4.10)
\mathfrak{L}	label configuration $\mathbb{T}^2 \rightarrow [20]$ used in experiments (5.27)

Symbols for discretized quantities

S	assignment matrix, spatial discretization of \mathfrak{S} (5.2)
L_Ω	Ω -induced discrete Laplacian (2.27)
P	assignment matrix, starting point for S flow (2.24), (2.29), spatial discretization of \mathfrak{P}
h	discretization of h (5.3)
L_h	discretization of Δ_h (5.8)
D	discretization of ∂_μ (5.11)
O	discretization of \mathcal{O} (5.5)

Calligraphic Symbols

\mathcal{H}	Gaussian mean curvature
\mathcal{N}	surface normal vector
\mathcal{W}_c^N	assignment manifold (2.22)
\mathcal{G}	graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
\mathcal{V}	graph vertices
\mathcal{E}	graph edges
\mathcal{J}	S flow potential (2.23)
\mathcal{O}	operator mapping states to metrics (3.18), (3.20), (3.26), (4.4), (5.30), (5.33) (σ -flow), (σ - α flow), (σ - φ -flow), (σ - φ - α flow)
\mathcal{C}	class of functions parametrized by a neural network (5.32)
\mathcal{B}	matrix function (4.29)
\mathcal{F}	metric space

We deliberately overload some symbols when there is no danger of confusion due to the context.

A.2 Implementation Details

Numerics. For assessing the convergence behavior of the sigma flow (Section 5.4), we numerically solved the ODE problem corresponding to the semi-discrete sigma flow (5.12) using a geometric Runge-Kutta scheme of order five with adaptive step size implemented via the Python libraries DiffraX [Kid21] and Jax [BFH⁺18], and with the standard 5-point stencil for discretizing the flat Laplacian Δ on \mathbb{T}^2 . The Adabelief first-order optimizer [ZTD⁺20] was employed with step size 0.01.

For the computationally more expensive experiments reported in Sections 5.5 and 5.6, we used the simplest geometric Runge-Kutta scheme, i.e. geometric Euler integration, to speed up the optimization, with end time $T = 2$, mass parameter $m^2 = 4$ and a constant step size of 0.2. In the first experiment, we used the Adabelief first order optimizer [ZTD⁺20] with step size 0.01. Here the step size of Adabelief was reduced to 10^{-4} .

Noisy labelings. The corrupted labelings for the experiments in Section 5.6 were created by first applying label smoothing with smoothing factor 0.8, followed by applying the logarithm, projection to the unit sphere, addition of noise sampled from the standard normal distribution $\mathcal{N}(0, 0.2)$, projecting again to the unit sphere and applying softmax. For visualization, we rounded pixelwise to the label with highest probability and applied a color code.

Parametrization of the metric h . We specify the parametrization of the positive definite 2×2 matrices $h(t)_a$ (cf. (5.4)). For simplicity, we drop the subscript a indicating the dependency on the grid points that are used to discretize the domain manifold M . We reuse below the symbol a with a different meaning confined to this section.

The three degrees of freedom of the metric tensor are given by $x, y, z \in \mathbb{R}$. The inverse metric tensor bounded from below by $0.01\mathbb{I}_2$ has the form

$$h^{-1} = \frac{1}{v} \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \quad (\text{A.1})$$

where we set

$$\lambda = \sigma(x) + 0.01, \quad \alpha = \frac{\pi}{2} \tanh(y), \quad v = \sigma(z) + 0.01 \quad (\text{A.2})$$

with the sigmoid function

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (\text{A.3})$$

and the matrix entries

$$a = \lambda + \sin^2(\alpha)\Delta, \quad b = \lambda \sin(\alpha) \cos(\alpha), \quad c = \lambda + \cos^2(\alpha)\Delta, \quad \Delta = \frac{1}{\lambda} - \lambda. \quad (\text{A.4})$$

This parametrization guarantees that the resulting matrix h^{-1} is symmetric positive definite and its smallest eigenvalue is not smaller than 0.01. For the experiments reported in Section 5.5, we applied this function to a parameter matrix of size $\mathbb{R}^{N \times 3}$, described below as part of the neural network architecture, along the last dimension to obtain a metric.

Neural network. The neural network employed for the learning experiments of Section 5.6 contained a single convolutional layer with kernel size 15×15 and 64 filters padded so as to preserve the number of vertices in the discretized tori. Following the convolutional layer, we applied a multilayer perceptron with 3 hidden layers of dimensions 32, 16, 4 and output dimension 3 at each pixel separately. Using the 3 dimensional output vectors, the metric h^{-1} was constructed by applying of the function $\mathbb{R}^3 \rightarrow \mathbb{R}^{2 \times 2}$ described above.

References

- [AGLM93] L. Alvarez, F. Guichard, P. L. Lions, and J. M. Morel, *Axioms and Fundamental Equations of Image Processing*, Arch. Rat. Mech. Anal. **123** (1993), 199–257.
- [AJLS17] N. Ay, J. Jost, H. V. Lê, and L. Schwachhöfer, *Information Geometry*, Springer, 2017.
- [AN00] S.-I. Amari and H. Nagaoka, *Methods of Information Geometry*, American Mathematical Society, 2000.
- [ÅPSS17] F. Åström, S. Petra, B. Schmitzer, and C. Schnörr, *Image Labeling by Assignment*, Journal of Mathematical Imaging and Vision **58** (2017), no. 2, 211–238.
- [Bat11] T. Batard, *Heat Equations on Vector Bundles – Application to Color Image Regularization*, Journal of Mathematical Imaging and Vision **41** (2011), no. 1, 59–85.
- [BBL⁺17] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, *Geometric Deep Learning: Going beyond Euclidean data*, IEEE Signal Processing Magazine **34** (2017), no. 4, 18–42.
- [BCA⁺24] B. Boll, J. Cassel, P. Albers, S. Petra, and C. Schnörr, *A Geometric Embedding Approach to Multiple Games and Multiple Populations*, preprint arXiv:2401.05918 (2024).
- [BCG⁺18] U. V. Boscain, R. Chertovskih, J.-P. Gauthier, D. Prandi, and A. Remizov, *Highly Corrupted Image inpainting Through Hypoelliptic Diffusion*, J. Math. Imag. Vision **60** (2018), no. 8, 1231–1245.
- [BCM06] A. Buades, B. Coll, and J.-M. Morel, *Neighborhood Filter and PDEs*, Numer. Math. **105** (2006), 1–34.
- [BF16] A.L. Bertozzi and A. Flenner, *Diffuse Interface Models on Graphs for Classification of High Dimensional Data*, SIAM Review **58** (2016), no. 2, 293–328.
- [BFH⁺18] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang, *JAX: composable transformations of Python+NumPy programs*, 2018.
- [BGAPS24] B. Boll, D. Gonzalez-Alvarado, S. Petra, and C. Schnörr, *Generative Assignment Flows for Representing and Learning Joint Distributions of Discrete Data*, preprint arXiv:2406.04527 (2024).
- [BKP10] K. Bredies, K. Kunisch, and T. Pock, *Total Generalized Variation*, SIAM Journal on Imaging Sciences **3** (2010), no. 3, 492–526.
- [BS14] T. Batard and N. Sochen, *A Class of Generalized Laplacians on Vector Bundles Devoted to Multi-Channel Image Processing*, Journal of Mathematical Imaging and Vision **48** (2014), no. 3, 517–543.
- [BZPS23] B. Boll, A. Zeilmann, S. Petra, and C. Schnörr, *Self-Certifying Classification by Linearized Deep Assignment*, PAMM: Proc. Appl. Math. Mech. **23** (2023), no. 1, e202200169.
- [CCN15] V. Caselles, A. Chambolle, and M. Novaga, *Total Variation in Imaging*, Handbook of Mathematical Methods in Imaging (O. Scherzer, ed.), Springer, 2015, pp. 1455–1499.
- [CCP12] A. Chambolle, D. Cremers, and T. Pock, *A Convex Approach to Minimal Partitions*, SIAM J. Imag. Sci. **5** (2012), no. 4, 1113–1158.
- [CFSS16] G. Citti, B. Franceschiello, G. Sanguinetti, and A. Sarti, *Sub-Riemannian Mean Curvature Flow for Image Processing*, SIAM J. Imag. Sci. **9** (2016), no. 1, 212–237.
- [Cha84] I. Chavel, *Eigenvalues in Riemannian Geometry*, Pure and Applied Mathematics, no. 115, Academic Press, Orlando, 1984.

- [Cha04] A. Chambolle, *An Algorithm for Total Variation Minimization and Applications*, Journal of Mathematical Imaging and Vision **20** (2004), no. 1, 89–97.
- [CRBD18] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, *Neural Ordinary Differential Equations*, NIPS (2018).
- [CRE⁺21] B. Chamberlain, J. Rowbottom, D. Eynard, F. Di Giovanni, X. Dong, and M. Bronstein, *Beltrami Flow and Neural Diffusion on Graphs*, Advances in Neural Information Processing Systems **34** (2021), 1594–1609.
- [DHLM05] M. Desbrun, A. N. Hirani, M. Leok, and J. E. Marsden, *Discrete Exterior Calculus*, preprint arXiv:math/0508341 (2005).
- [DMSC16] J. Duran, M. Moeller, C. Sbert, and D. Cremers, *Collaborative Total Variation: A General Framework for Vectorial TV Models*, SIAM J. Imaging Sciences **9** (2016), no. 1, 116–151.
- [EKV23] M. Elad, B. Kawar, and G. Vaksman, *Image Denoising: The Deep Learning Revolution and Beyond—A Survey Paper*, SIAM J. Imaging Sciences **16** (2023), no. 3, 1033–1855.
- [FM19] P. M. N. Feehan and M. Maridakis, *Lojasiewicz–Simon Gradient Inequalities for the Harmonic Map Energy Function*, preprint arXiv:1903.01953v1 (2019).
- [FMN16] C. Fefferman, S. Mitter, and H. Narayanan, *Testing the Manifold Hypothesis*, J. Amer. Math. Soc. **29** (2016), no. 4, 983–1049.
- [Gar13] H. Garcke, *Curvature Driven Interface Evolution*, Jahresberichte der Deutschen Mathematiker-Vereinigung **115** (2013), no. 2, 63–100.
- [GLM20] J. Gaster, B. Loustau, and L. Monsaingeon, *Computing Discrete Equivariant Harmonic Maps*, preprint arXiv:1810.11932 (2020).
- [GLPR23] B. Geshkovski, C. Letrouit, Y. Polyanskiy, and P. Rigollet, *A Mathematical Perspective on Transformers*, preprint arXiv:2312.10794 (2023).
- [GY02] X. Gu and S.-T. Yau, *Computing Conformal Structures of Surfaces*, Commun. Information and Systems **2** (2002), no. 2, 121–146.
- [Ham17] M. J. D. Hamilton, *Mathematical Gauge Theory*, Springer, 2017.
- [HJ13] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed., Cambridge University Press, 2013.
- [HJLZ19] X. Han, J. Jost, L. Liu, and L. Zhao, *Global Existence of the Harmonic Map Heat Flow into Lorentzian Manifolds*, J. Math. Pures Appl. **130** (2019), 130–156.
- [HKK⁺03] K. Hori, S. Katz, A. Klemm, R. Pandharipande, R. Thomas, C. Vafa, R. Vakil, and E. Zaslov, *Mirror Symmetry*, Amer. Math. Soc., 2003.
- [HKW77] S. Hildebrandt, H. Kaul, and K.-O. Widman, *An Existence Theorem for Harmonic Mappings of Riemannian Manifolds*, Acta Math. **138** (1977), no. 1, 1–16.
- [HS98] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press, 1998.
- [HSLG23] A. Han, D. Shi, L. Lin, and J. Gao, *From Continuous Dynamics to Graph Neural Networks: Neural Diffusion and Beyond*, preprint arXiv:2310.10121 (2023).
- [HSPS21] R. Hühnerbein, F. Savarino, S. Petra, and C. Schnörr, *Learning Adaptive Regularization for Image Labeling Using Geometric Assignment*, J. Math. Imaging Vision **63** (2021), 186–215.
- [Hua06] S.-Z. Huang, *Gradient Inequalities*, Amer. Math. Soc., 2006.

- [HW08] F. Hélein and J. C. Wood, *Harmonic Maps*, Handbook of Global Analysis, Elsevier, 2008, pp. 417–491.
- [Jos97] J. Jost, *Nonpositive Curvature: Geometric and Analytic Aspects*, Springer Basel AG, 1997.
- [Jos17] ———, *Riemannian Geometry and Geometric Analysis*, Springer, 2017.
- [JS09] J. Jost and F. M. Simsir, *Affine Harmonic Maps*, *Analysis* **29** (2009), no. 2, 185–197.
- [Kas89] R. E. Kass, *The Geometry of Asymptotic Inference*, *Statistical Science* **4** (1989), no. 3, 188–219.
- [Kic08] S. Kichenassamy, *The Perona-Malik Method as an Edge Pruning Algorithm*, *J. Math. Imaging Vision* **30** (2008), 209–219.
- [Kid21] P. Kidger, *On Neural Differential Equations*, Ph.D. thesis, University of Oxford, 2021.
- [KMS00] R. Kimmel, R. Malladi, and N. Sochen, *Images as Embedded Maps and Minimal Surfaces: Movies, Color, Texture, and Volumetric Medical Images*, *International Journal of Computer Vision* **39** (2000), 111–129.
- [Lim20] L.-H. Lim, *Hodge Laplacians on Graphs*, *SIAM Review* **62** (2020), no. 3, 685–715.
- [LRMU15] S. Lefkimmiatis, A. Roussos, P. Maragos, and M. Unser, *Structure Tensor Total Variation*, *SIAM J. Imaging Sciences* **8** (2015), no. 2, 1090–1122.
- [LS11] J. Lellmann and C. Schnörr, *Continuous Multiclass Labeling Approaches and Algorithms*, *SIAM J. Imag. Sci.* **4** (2011), no. 4, 1049–1096.
- [MB06] M. Moakher and P. G. Batchelor, *Symmetric Positive-Definite Matrices: From Geometry to Applications and Visualization*, *Visualization and Processing of Tensor Fields* (J. Weickert and H. Hagen, eds.), Springer, Berlin, Heidelberg, 2006, pp. 285–298.
- [MBC15] J.-M. Morel, A. Buades, and T. Coll, *Local Smoothing Neighborhood Filters*, *Handbook of Mathematical Methods in Imaging* (O. Scherzer, ed.), Springer, 2015, pp. 1599–1643.
- [MBO94] B. Merriman, J.K. Bence, and S.J. Osher, *Motion of Multiple Junctions: A Level Set Approach*, *J. Comp. Physics* **112** (1994), no. 2, 334–363.
- [MH80] D. Marr and E. Hildreth, *Theory of Edge Detection*, *Proc. Royal Soc. B* **207** (1980), 187–217.
- [MOA11] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, 2nd ed., Springer, 2011.
- [Nis02] S. Nishikawa, *Variational Problems in Geometry*, Amer. Math. Soc., 2002.
- [PC17] J. Pang and G. Cheung, *Graph Laplacian Regularization for Image Denoising: Analysis in the Continuous Domain*, *IEEE Trans. Image Proc.* **26** (2017), no. 4, 1770–1785.
- [PM90] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12** (1990), no. 7, 629–639.
- [Pol81] A. M. Polyakov, *Quantum Geometry of Bosonic Strings*, *Physics Letters B* **103** (1981), no. 3, 207–210.
- [Pol98] J. Polchinski, *String Theory: An Introduction to the Bosonic String*, vol. 1, Cambridge University Press, 1998.
- [Pro16] E. Provenzi, *A Differential Geometry Model for the Perceived Colors Space*, *Int. J. Geom. Meth. Modern Physics* **13** (2016), no. 8, 1630008.

- [Res74] H.L. Resnikoff, *Differential Geometry and Color Perception*, J. Math. Biology **1** (1974), 97–131.
- [ROF92] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear Total Variation Based Noise Removal Algorithms*, Physica D **60** (1992), 259–268.
- [Ros97] S. Rosenberg, *The Laplacian on a Riemannian Manifold*, Cambridge Univ. Press, 1997.
- [SBS23] D. Sitenko, B. Boll, and C. Schnörr, *A Nonlocal Graph-PDE and Higher-Order Geometric Integration for Image Labeling*, SIAM J. Imaging Sciences **16** (2023), no. 1, 501–567.
- [SCB⁺23] J. Schwarz, J. Cassel, B. Boll, M. Gärtner, P. Albers, and C. Schnörr, *Quantum State Assignment Flows*, Entropy **25** (2023), no. 9, 1253.
- [Sch15] O. Scherzer (ed.), *Handbook of Mathematical Methods in Imaging*, Springer, 2015.
- [Sch20] C. Schnörr, *Assignment Flows*, Handbook of Variational Methods for Nonlinear Geometric Data (P. Grohs, M. Holler, and A. Weinmann, eds.), Springer, 2020, pp. 235–260.
- [Shi07] H. Shima, *The Geometry of Hessian Structures*, World Scientific, 2007.
- [Sim83] L. Simon, *Asymptotics for a Class of Non-Linear Evolution Equations, with Applications to Geometric Problems*, Ann. Mathematics **118** (1983), no. 3, 525–571.
- [SKM98] N. Sochen, R. Kimmel, and R. Malladi, *A General Framework for Low Level Vision*, IEEE transactions on image processing: a publication of the IEEE Signal Processing Society **7** (1998), no. 3, 310–318.
- [Sol85] B. Solomon, *Harmonic Maps to Spheres*, Journal of Differential Geometry **21** (1985), no. 2, 151–162.
- [SPBD23] B. Smets, J. Portegies, E. Bekkers, and R. Duits, *PDE-based Group Equivariant Convolutional Neural Networks*, Journal of Mathematical Imaging and Vision **65** (2023), no. 1, 209–239.
- [SPSOD21] B. M. N. Smets, J. Portegies, E. St-Onge, and R. Duits, *Total Variation and Mean Curvature PDEs on the Homogeneous Space of Positions and Orientations*, J. Math. Imaging Vision **63** (2021), 237–262.
- [SS21] F. Savarino and C. Schnörr, *Continuous-Domain Assignment Flows*, Europ. J. Appl. Math. **32** (2021), no. 3, 570–597.
- [Uoh14] K. Uohashi, *Harmonic Maps Relative to α -Connections*, Geometric Theory of Information, Springer, 2014, pp. 81–96.
- [vGGOB14] Y. van Gennip, N. Guillen, B. Osting, and A. L. Bertozzi, *Mean Curvature, Threshold Dynamics, and Phase Field Theory on Finite Graphs*, Milan Journal of Mathematics **82** (2014), no. 1, 3–65.
- [VSP⁺17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, and L. Kaiser, *Attention Is All You Need*, NIPS, 2017.
- [War08] M. Wardetzky, *Convergence of the Cotangent Formula: An Overview*, Discrete Differential Geometry, Oberwolfach Seminars, vol. 38, Birkhäuser, 2008, pp. 275–286.
- [Wei98] J. Weickert, *Anisotropic Diffusion in Image Processing*, B.G. Teubner Verlag, 1998.
- [WY23] T. Wu and S.-T. Yau, *Computing Harmonic Maps and Conformal Maps on Point Clouds*, J. Comp. Math. **41** (2023), 879–908.
- [ZSPS20] A. Zeilmann, F. Savarino, S. Petra, and C. Schnörr, *Geometric Numerical Integration of the Assignment Flow*, Inverse Problems **36** (2020), no. 3, 034004.

- [ZTD⁺20] Juntang Zhuang, Tommy Tang, Yifan Ding, Sekhar C Tatikonda, Nicha Dvornek, Xenophon Papademetris, and James Duncan, *Adabelief optimizer: Adapting stepsizes by the belief in observed gradients*, Advances in Neural Information Processing Systems **33** (2020).
- [ZZPS20a] A. Zern, M. Zisler, S. Petra, and C. Schnörr, *Unsupervised Assignment Flow: Label Learning on Feature Manifolds by Spatially Regularized Geometric Assignment*, Journal of Mathematical Imaging and Vision **62** (2020), no. 6–7, 982–1006.
- [ZZPS20b] M. Zisler, A. Zern, S. Petra, and C. Schnörr, *Self-Assignment Flows for Unsupervised Data Labeling on Graphs*, SIAM Journal on Imaging Sciences **13** (2020), no. 3, 1113–1156.
- [ZZS22] A. Zern, A. Zeilmann, and C. Schnörr, *Assignment Flows for Data Labeling on Graphs: Convergence and Stability*, Information Geometry **5** (2022), 355–404.