Euro. Jnl of Applied Mathematics (2021), *vol.* 32, *pp.* 570–597
doi:10.1017/S0956792520000273

Continuous-domain assignment flows

F. SAVARINO and C. SCHNÖRR

Image and Pattern Analysis Group, Heidelberg University, Heidelberg, Germany
emails: fabrizio.savarino@iwr.uni-heidelberg.de; schnoerr@math.uni-heidelberg.de
URL: https://ipa.math.uni-heidelberg.de

(*Received 2 October 2019*; *revised 10 April 2020*; *accepted 1 August 2020*; *first published online 1 September 2020*)

Assignment flows denote a class of dynamical models for contextual data labelling (classification) on graphs. We derive a novel parametrisation of assignment flows that reveals how the underlying information geometry induces two processes for assignment regularisation and for gradually enforcing unambiguous decisions, respectively, that seamlessly interact when solving for the flow. Our result enables to characterise the dominant part of the assignment flow as a Riemannian gradient flow with respect to the underlying information geometry. We consider a continuous-domain formulation of the corresponding potential and develop a novel algorithm in terms of solving a sequence of linear elliptic partial differential equations (PDEs) subject to a simple convex constraint. Our result provides a basis for addressing learning problems by controlling such PDEs in future work.

Key words: Image labelling, image segmentation, information geometry, replicator equation, evolutionary dynamics, assignment flow

2020 Mathematics Subject Classification: 35R02, 62M45, 68U10 (Primary); 49J40, 91A22 (Secondary)

1 Introduction

Deep networks are omnipresent in many disciplines due to their unprecedented predictive power and the availability of software for training that is easy to use. However, this rapid development during recent years has not improved our mathematical understanding in the same way, so far [14]. The 'black box' behaviour of deep networks and systematic failures [5], the lack of performance guarantees and reproducibility of results raise doubts if a purely data-driven approach can deliver the high expectations of some of its most passionate proponents [11]. 'Mathematics of deep networks', therefore, has become a focal point of research.

Based on residual networks, introduced in [16], attempts to understand deep network architectures as discretised realisations of dynamical systems mathematically substantiated and promoted by, for example, [15, 30], have become a fruitful line of research. Adopting this viewpoint, we introduced a dynamical system – called *assignent flow* – for contextual data classification and image labelling on graphs [6]. We refer to [27] for a review of recent work including parameter estimation (learning) [18], adaption of data prototypes during assignment [34] and learning prototypes from low-rank data representations and self-assignment [37]. Convergence and stability of the assignment flow have been studied by [33].

Two key properties of the assignment flow are *smoothness* and *gradual enforcement* of unambigious classification in a *single* process, solely induced by adopting an elementary statistical manifold as state space that is natural for classification tasks, and the corresponding information geometry [2]. This differs from traditional *variational* approaches to image labelling [21, 10] that

enjoy convexity but are inherently non-smooth and require post-processing to achieve unambigous decisions. In addition, hierarchical combinations of such variational approaches lead to *non-smooth* non-convex multi-level optimisation problems. By contrast, the corresponding extensions of the assignment flow, such as simultaneous adaption of feature prototypes [34], yield more complex but still *smooth* dynamical systems, the flows of which can be conveniently computed using geometric integration.

The assignment flow combines by *composition* (rather than by addition) separate local processes at each vertex of the underlying graph and non-local regularisation. Each local process for label assignment is governed by an ordinary differential equation (ODE), the *replicator equation* [17, 26], whereas regularisation is accomplished by non-local *geometric* averaging of the evolving assignments. It is well known [17] that if the affinity measure which defines the replicator equation and hence governs label selection can be derived as gradient of a potential, then the replicator equation is just the corresponding Riemannian gradient flow induced by the Fisher–Rao metric. The geometric regularisation of assignments performed by the assignment flow yields an affinity measure for which the (non-)existence of a corresponding potential is not immediate, however.

Contribution and Organisation. The objective of this paper is to clarify this situation. After collecting background material in Section 2, we prove that no potential exists that enables to characterise the assignment flow as Riemannian gradient flow (Section 3.1). Next, we provide a novel parametrisation of the assignment flow by separating a dominant component of the flow, called *S-flow*, that completely determines the remaining component and hence essentially characterises the assignment flow (Section 3.2). The *S*-flow *does* correspond to a potential, under an additional symmetry assumption with respect to the weights that parametrise the regularisation properties of the assignment flow through (weighted) geometric averaging. Based on this result, convergence and stability of the assignment flow have been studied recently [33]. This potential can be decomposed into two components that make explicit the two interacting processes mentioned above: regularisation of label assignments and gradually enforcing unambigous decisions. We point out again that this is a direct consequence of the 'spherical geometry' (positive curvature) underlying the assignment flow.

Based on this result, we consider the corresponding *continuous-domain variational* formulation in Section 4. We prove well-posedness which is not immediate due to non-convexity, and we propose an algorithm that computes a locally optimal assignment by solving a sequence of simple *linear* PDEs, with changing right-hand side and subject to a simple convex constraint. A numerical example demonstrates that our PDE-based approach reproduces results obtained with solving the original formulation of the assignment flow using completely different numerical techniques [35]. We hope that the *simplicity* of our PDE-approach and the direct connection to a smooth geometric setting will stimulate future work on learning, from an optimal control point of view [31, 22]. We conclude by a formal derivation of a PDE that characterises global minimisers of the non-convex objective function (Section 4.4) and by outlining future research in Section 5.

2 Preliminaries

2.1 Basic notation

We denote the standard basis of $\mathbb{R}^n$ by

$$\mathcal{B}_n := \{e_1, \ldots, e_n\}. \tag{2.1}$$

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

$|\cdot|$ applied to a finite set denotes its cardinality, that is, $|\mathcal{B}_n| = n$. We set $[n] = \{1, 2, \dots, n\}$ for $n \in \mathbb{N}$ and $\mathbb{1}_n = (1, 1, \dots, 1)^\top \in \mathbb{R}^n$. The symbols

$$\mathcal{I} = [n], \qquad \mathcal{J} = [c], \qquad n, c \in \mathbb{N} \tag{2.2}$$

will specifically index data points and classes (labels), respectively. $\|\cdot\|$ denotes the Euclidean vector norm and the Frobenius matrix norm induced by the inner product $\|A\| = \langle A, A\rangle^{1/2} = \mathrm{tr}(A^\top A)^{1/2}$. All other norms will be indicated by a corresponding subscript. For a given matrix $A \in \mathbb{R}^{n\times c}$, A_i, $i \in [n]$ denote the row vectors, A^j, $j \in [c]$ denote the column vectors and $A^\top \in \mathbb{R}^{c\times n}$ the transpose matrix. $\mathbb{R}^n_+$ is the set of vectors in $\mathbb{R}^n$ with nonnegative entries and

$$\Delta_n = \{p \in \mathbb{R}^n_+ \colon \langle \mathbb{1}_n, p\rangle = 1\} \tag{2.3}$$

denotes the probability simplex. There will be no danger to confuse it with the Laplacian differential operator Δ that we use without subscript. For strictly positive vectors $p > 0$, we efficiently denote componentwise subdivision by $\frac{v}{p}$. Likewise, we set $pv = (p_1 v_1, \dots, p_n v_n)^\top$. The exponential *function* applies componentwise to vectors (and similarly for log) and will always be denoted by $e^v = (e^{v_1}, \dots, e^{v_n})^\top$, in order not to confuse it with the exponential *maps* (2.19).

Strong and weak convergence of a sequence (f_n) is written as $f_n \to f$ and $f_n \rightharpoonup f$, respectively. ψ_S denotes the indicator function of some set S: $\psi_S(i) = 1$ if $i \in S$ and $\psi_S(i) = 0$ otherwise. δ_C denotes the indicator function from the optimisation point of view: $\delta_C(f) = 0$ if $f \in C$ and $\delta_C(f) = +\infty$ otherwise.

2.2 Assignment manifold and flow

We sketch the assignment flow as introduced by [6] and refer to the recent survey [27] for more background and a review of recent related work.

2.2.1 *Assignment manifold*

Suppose $\mathcal{I}$ is a set of $n := |\mathcal{I}|$ nodes and let

$$f \colon \mathcal{I} \to \mathcal{F}, \quad i \mapsto f_i \in \mathcal{F} \tag{2.4}$$

represent some given data with values in a metric space $(\mathcal{F}, d_\mathcal{F})$. Further, assume that a set of predefined prototypes

$$\mathcal{F}_* = \{f_j^* \in \mathcal{F} \colon j \in \mathcal{J}\}, \qquad |\mathcal{J}| = c, \tag{2.5}$$

also called *labels*, is given. *Data labelling* denotes the task of assigning to each node $i \in \mathcal{I}$ a label in $\mathcal{F}_*$, depending on the data f in some spatial context

$$\mathcal{I} \to \mathcal{F}_*, \qquad i \mapsto f^*_{j_i}. \tag{2.6}$$

If the given data are an image on a grid graph whose nodes and edges index the pixels and a neighbourhood structure, then the labelling task is also referred to as *image labelling*, see Figure 1 for an illustration.

In order to device a mathematical formulation for this dependence on the data, the set $\mathcal{I}$ is assumed to form the vertex set of an undirected graph $\mathcal{G} = (\mathcal{I}, \mathcal{E})$ which defines a relation $\mathcal{E} \subset \mathcal{I} \times \mathcal{I}$ and neighbourhoods

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

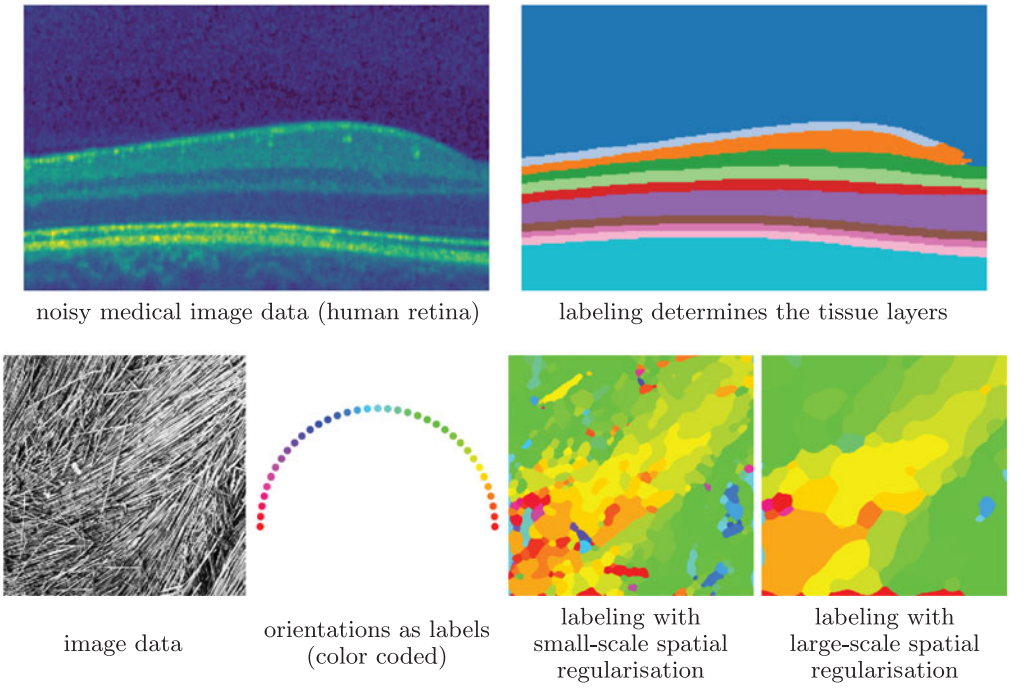

FIGURE 1. Two examples of image labelling. Top, from left to right: noisy medical image data of retinal tissue obtained by a non-invasive imaging method. The labelling task is to segment the different layers of tissue. Bottom, from left to right: an image depicting a texture with various orientations. The (colour coded) labels in this case are the different orientations of $0, 5, \ldots, 175$ degrees corresponding to a bank of orientation-selective Gabor filters that is applied to the image data in a pre-processing step. The task is to estimate the dominant orientation at every pixel depending on the spatial context. Labelling with small-scale spatial regularisation resolves more details but is also susceptible to noise, whereas large-scale regularisation labels the dominant orientations in a robust way.

$$\mathcal{N}_i := \{k \in \mathcal{I} \colon ik \in \mathcal{E}\} \cup \{i\}, \tag{2.7}$$

where ik is a shorthand for the unordered pair (edge) $(i, k) = (k, i)$. We require these neighbourhoods to satisfy the relations

$$k \in \mathcal{N}_i \quad \Leftrightarrow \quad i \in \mathcal{N}_k, \qquad \forall i, k \in \mathcal{I}. \tag{2.8}$$

The assignments (labelling) (2.6) are represented by matrices in the set

$$\mathcal{W}_* := \{W \in \{0, 1\}^{n \times c} \colon W \mathbb{1}_c = \mathbb{1}_n\} \tag{2.9}$$

with unit vectors W_i, $i \in \mathcal{I}$, called *assignment vectors*, as row vectors. These assignment vectors are computed by numerically integrating the assignment flow below (2.32), in the following elementary geometric setting. The integrality constraint of (2.9) is relaxed and vectors

$$W_i = (W_{i1}, \ldots, W_{ic})^\top \in \mathcal{S}, \quad i \in \mathcal{I}, \tag{2.10}$$

that we still call *assignment vectors*, are considered on the elementary Riemannian manifold

$$(\mathcal{S}, g), \qquad \mathcal{S} := \{p \in \Delta_c \colon p > 0\}, \tag{2.11}$$

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

with

$$\mathbb{1}_{\mathcal{S}} := \frac{1}{c}\mathbb{1}_c \in \mathcal{S}, \qquad (\textit{barycenter}) \tag{2.12}$$

(constant) tangent space

$$T_0 := \{v \in \mathbb{R}^c : \langle \mathbb{1}_c, v\rangle = 0\}, \tag{2.13}$$

and tangent bundle $T\mathcal{S} = \mathcal{S} \times T_0$, orthogonal projection

$$\Pi_0 : \mathbb{R}^c \to T_0, \qquad \Pi_0 := I - \mathbb{1}_{\mathcal{S}}\mathbb{1}_c^\top, \tag{2.14}$$

and the Fisher–Rao metric

$$g_p(u, v) := \sum_{j\in\mathcal{J}} \frac{u^j v^j}{p^j}, \quad p \in \mathcal{S}, \quad u, v \in T_0. \tag{2.15}$$

For $p \in \mathcal{S}$, the *replicator operator* is the linear map

$$R_p : \mathbb{R}^c \to T_0, \qquad R_p := \operatorname{Diag}(p) - pp^\top, \tag{2.16}$$

satisfying

$$R_p = R_p\Pi_0 = \Pi_0 R_p \quad \text{and} \quad R_{\mathbb{1}_{\mathcal{S}}} = \frac{1}{c}\Pi_0. \tag{2.17}$$

For a vector field $F : \mathcal{S} \to \mathbb{R}^c$, representing an affinity measure, the corresponding replicator equation is given by

$$\dot{p}_k = p_k(F_k(p) - \mathbb{E}_p[F]) = p_k F_k(p) - \langle p, F(p)\rangle p_k = \big(R_p F(p)\big)_k \quad \text{for all } k \in [c]. \tag{2.18}$$

Thus, the replicator operator R_p can be seen as the parameterisation of the *replicator equation* [17] given the vector field F.

The exponential map and its inverse are defined as

$$\operatorname{Exp} : \mathcal{S} \times T_0 \to \mathcal{S}, \qquad (p, v) \mapsto \operatorname{Exp}_p(v) := \frac{pe^{\frac{v}{p}}}{\langle p, e^{\frac{v}{p}}\rangle}, \tag{2.19a}$$

$$\operatorname{Exp}_p^{-1} : \mathcal{S} \to T_0, \qquad q \mapsto \operatorname{Exp}_p^{-1}(q) = R_p \log \frac{q}{p}, \tag{2.19b}$$

and using R_p, the *lifting map* and its inverse are given by

$$\exp_p : T_0 \to \mathcal{S}, \qquad \exp_p := \operatorname{Exp}_p \circ R_p, \tag{2.19c}$$

$$\exp_p^{-1} : \mathcal{S} \to T_0, \qquad \exp_p^{-1}(q) = \Pi_0 \log \frac{q}{p}. \tag{2.19d}$$

Applying the map $\exp_p$ to a vector in $\mathbb{R}^c = T_0 \oplus \mathbb{R}\mathbb{1}$ does not depend on the constant component of the argument, due to (2.17).

Remark 2.1 *(1) The map* Exp *is* ***not*** *the exponential map of the Riemannian connection but corresponds to the e-connection of information geometry [2]. Accordingly, the affine geodesics* (2.19a) *are not length-minimising. However, they provide a close approximation [6, Proposition 3] and are more convenient for numerical computations.*

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

(2) For $x \in \mathbb{R}^c$ and $p \in \mathcal{S}$, the lifting map can alternatively be characterised by

$$\exp_p(-x) = \operatorname{argmin}_{q \in \Delta} \big\{ \langle q, x \rangle + \mathrm{KL}(q, p) \big\}, \tag{2.20}$$

a projected gradient step [7] with respect to the KL *divergence.*

The *assignment manifold* is defined as

$$(\mathcal{W}, g), \qquad \mathcal{W} := \mathcal{S} \times \cdots \times \mathcal{S}. \qquad (n = |\mathcal{I}| \text{ factors}) \tag{2.21}$$

We identify $\mathcal{W}$ with the embedding into $\mathbb{R}^{n \times c}$

$$\mathcal{W} = \{ W \in \mathbb{R}^{n \times c} \colon W \mathbb{1}_c = \mathbb{1}_n \text{ and } W_{ij} > 0 \text{ for all } i \in [n], j \in [c] \}. \tag{2.22}$$

Thus, points $W \in \mathcal{W}$ are row-stochastic matrices $W \in \mathbb{R}^{n \times c}$ with row vectors $W_i \in \mathcal{S}$, $i \in \mathcal{I}$ that represent the assignments (2.6) for every $i \in \mathcal{I}$. We set

$$\mathcal{T}_0 := T_0 \times \cdots \times T_0 \qquad (n = |\mathcal{I}| \text{ factors}). \tag{2.23}$$

Due to (2.22), the tangent space $\mathcal{T}_0$ can be identified with

$$\mathcal{T}_0 = \{ V \in \mathbb{R}^{n \times c} \colon V \mathbb{1}_c = 0 \}. \tag{2.24}$$

Thus, $V_i \in T_0$ for all row vectors of $V \in \mathbb{R}^{n \times c}$ and $i \in \mathcal{I}$. All mappings defined above factorise in a natural way and apply row-wise, for example, $\operatorname{Exp}_W = (\operatorname{Exp}_{W_1}, \ldots, \operatorname{Exp}_{W_n})$.

2.2.2 *Assignment flow*

Based on (2.4) and (2.5), the distance vector field

$$D_{\mathcal{F};i} := \big(d_{\mathcal{F}}(f_i, f_1^*), \ldots, d_{\mathcal{F}}(f_i, f_c^*) \big)^\top, \qquad i \in \mathcal{I}, \tag{2.25}$$

is well defined. These vectors are collected as row vectors of the *distance matrix*

$$D_{\mathcal{F}} \in \mathbb{R}_+^{n \times c}. \tag{2.26}$$

The *likelihood map* and the *likelihood vectors*, respectively, are defined as

$$L_i \colon \mathcal{S} \to \mathcal{S}, \qquad L_i(W_i) := \exp_{W_i}\left(-\frac{1}{\rho} D_{\mathcal{F};i} \right) = \frac{W_i e^{-\frac{1}{\rho} D_{\mathcal{F};i}}}{\langle W_i, e^{-\frac{1}{\rho} D_{\mathcal{F};i}} \rangle}, \qquad i \in \mathcal{I}, \tag{2.27}$$

where the scaling parameter $\rho > 0$ is used for normalising the a priori unknown scale of the components of $D_{\mathcal{F};i}$ that depends on the specific application at hand. In this way, the distance information $D_{\mathcal{F};i}$ for each node $i \in \mathcal{I}$ is *lifted* onto the assignment manifold, depending on the current assignment W_i.

A key component of the assignment flow is the interaction of the likelihood vectors through *geometric* averaging within the local neighbourhoods (2.7). Specifically, using weights,

$$\omega_{ik} > 0, \quad \text{for all } k \in \mathcal{N}_i,\ i \in \mathcal{I} \quad \text{with} \quad \sum_{k \in \mathcal{N}_i} w_{ik} = 1, \tag{2.28}$$

the *similarity map* and the *similarity vectors*, respectively, are defined as

$$S_i \colon \mathcal{W} \to \mathcal{S}, \qquad S_i(W) := \operatorname{Exp}_{W_i}\left(\sum_{k \in \mathcal{N}_i} w_{ik} \operatorname{Exp}_{W_i}^{-1}\big(L_k(W_k)\big) \right), \qquad i \in \mathcal{I}. \tag{2.29}$$

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

If Exp_{W_i} were the exponential map of the Riemannian (Levi-Civita) connection, then the argument inside the brackets of the right-hand side would just be the negative Riemannian gradient with respect to W_i of the centre of mass objective function comprising the points L_k, $k \in \mathcal{N}_i$, that is, the weighted sum of the squared Riemannian distances between W_i and L_k [19, Lemma 6.9.4]. In view of Remark 2.1, this interpretation is only approximately true mathematically but still correct informally: $S_i(W)$ moves W_i towards the geometric mean of the likelihood vectors L_k, $k \in \mathcal{N}_i$. Since $\mathrm{Exp}_{W_i}(0) = W_i$, this mean is equal to W_i if the aforementioned gradient vanishes.

Remark 2.2 *An alternative interpretation of the similarity matrix is given in [6]. For this, denote the weighted geometric mean of the likelihood vectors in the neighbourhood $\mathcal{N}_i$ around $i \in \mathcal{I}$ by*

$$\mathrm{gm}^{\omega_i}\left\{L_k(W_k)\right\}_{k\in\mathcal{N}_i} := \prod_{k\in\mathcal{N}_i} \left(L_k(W_k)\right)^{\omega_{ik}}. \tag{2.30}$$

Then, the similarity matrix is equivalently characterised as the normalised geometric mean, which in turn is the weighted center of mass with respect to the KL *divergence on $\mathcal{S}$, that is,*

$$S_i(W) = \frac{\mathrm{gm}^{\omega_i}\{L_k(W_k)\}_{k\in\mathcal{N}_i}}{\left\langle \mathbb{1}_c, \mathrm{gm}^{\omega_i}\{L_k(W_k)\}_{k\in\mathcal{N}_i}\right\rangle} = \mathrm{argmin}_{p\in\Delta}\left\{ \sum_{k\in\mathcal{N}_i} \omega_{ik}\, \mathrm{KL}\left(p, L_k(W_k)\right)\right\}. \tag{2.31}$$

The *assignment flow* is induced by the locally coupled system of non-linear ODEs

$$\dot{W} = R_W S(W), \qquad W(0) = \mathbb{1}_{\mathcal{W}}, \tag{2.32a}$$

$$\dot{W}_i = R_{W_i} S_i(W), \qquad W_i(0) = \mathbb{1}_{\mathcal{S}}, \quad i \in \mathcal{I}, \tag{2.32b}$$

where $\mathbb{1}_{\mathcal{W}} \in \mathcal{W}$ denotes the barycentre of the assignment manifold (2.21). In view of (2.18), the assignment flow (2.32b) is a set of spatially coupled replicator equations with the similarity matrix $S(W)$ as the corresponding local affinity measure, that represents the dependence of assignments on the given data in a spatial context. The solution curve $W(t) \in \mathcal{W}$ is numerically computed by geometric integration [35] and determines a labelling $W(T) \in \mathcal{W}_*$ for sufficiently large T after a trivial rounding operation.

2.3 Functional analysis

We record background material that will be used in Section 4.

2.3.1 *Sobolev spaces*

We list few basic definitions and fix the corresponding notation [36, 1]. Throughout this section, $\Omega \subset \mathbb{R}^d$ denotes an open-bounded domain.

We denote the inner product and the norm of functions $f, g \in L^2(\Omega)$ by

$$(f,g)_\Omega = \int_\Omega fg dx, \qquad \|f\|_\Omega = (f,f)_\Omega^{1/2}. \tag{2.33}$$

Functions f_1 and f_2 are equivalent and identified whenever they merely differ pointwise on a Lebesque-negligible set of measure zero. f_1 and f_2 are then said to be equal a.e. (almost everywhere). $H^1(\Omega) = W^{1,2}(\Omega)$ denotes the Sobolev space of functions f with square-integrable

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

weak derivatives $D^\alpha f$ up to order one. $H^1(\Omega)$ is a Hilbert space with inner product and norm denoted by

$$(f,g)_{1;\Omega} = \sum_{|\alpha|\leq 1} (D^\alpha f, D^\alpha g)_\Omega, \qquad \|f\|_{1;\Omega} = \left(\sum_{|\alpha|\leq 1} \|D^\alpha f\|_\Omega^2 \right)^{1/2}. \tag{2.34}$$

Lemma 2.3 ([36, Corollary 2.1.9]) *If Ω is connected, $u \in H^1(\Omega)$ and $Du = 0$ a.e. on Ω, then u is equivalent to a constant function on Ω.*

The closure in $H^1(\Omega)$ of the set of test functions $C_c^\infty(\Omega)$ that are compactly supported on Ω is the Sobolev space

$$H_0^1(\Omega) = \overline{C_c^\infty(\Omega)} \subset H^1(\Omega). \tag{2.35}$$

It contains all functions in $H^1(\Omega)$ whose boundary values on $\partial\Omega$ (in the sense of traces) vanish. The space $H^1(\Omega;\mathbb{R}^c)$, $2 \leq c \in \mathbb{N}$ contains vector-valued functions f whose component functions f_i, $i \in [c]$ are in $H^1(\Omega)$. For notational efficiency, we denote the norm of $f \in H^1(\Omega;\mathbb{R}^c)$ by

$$\|f\|_{1;\Omega} = \left(\sum_{i\in[c]} \|f_i\|_{1;\Omega}^2 \right)^{1/2}, \tag{2.36}$$

as in the scalar case (2.34). It will be clear from the context if f is scalar- or vector-valued.

The compactness theorem of Rellich–Kondrakov [1, Theorem 5.3.3] says that the canonical embedding

$$H_0^1(\Omega) \hookrightarrow L^2(\Omega) \tag{2.37}$$

is compact, that is, every bounded subset in $H_0^1(\Omega)$ is relatively compact in $L^2(\Omega)$. This extends to the vector-valued case

$$H_0^1(\Omega;\mathbb{R}^c) \hookrightarrow L^2(\Omega;\mathbb{R}^c) \tag{2.38}$$

since $H_0^1(\Omega;\mathbb{R}^c)$ is isomorphic to $H_0^1(\Omega) \times \cdots \times H_0^1(\Omega)$ and likewise for $L^2(\Omega;\mathbb{R}^c)$. The dual space of $H_0^1(\Omega)$ is commonly denoted by $H^{-1}(\Omega) = \big(H_0^1(\Omega)\big)'$. Accordingly, we set $H^{-1}(\Omega;\mathbb{R}^c) = \big(H_0^1(\Omega;\mathbb{R}^c)\big)'$.

2.3.2 *Weak convergence properties, variational inequalities*

We list few further basic facts [32, Proposition 38.2], [1, Proposition 2.4.6].

Proposition 2.4 *The following assertions hold in a Banach space X.*

(a) *A closed convex subset $C \subset X$ is weakly closed, that is a sequence $(f_n)_{n\in\mathbb{N}} \subset C$ that weakly converges to f implies $f \in C$.*
(b) *If X is reflexive (in particular, if X is a Hilbert space), then every bounded sequence in X has a weakly convergent subsequence.*
(c) *If f_n weakly converges to f, then $(f_n)_{n\in\mathbb{N}}$ is bounded and*

$$\|f\|_X \leq \liminf_{n\to\infty} \|f_n\|_X. \tag{2.39}$$

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

The following theorem states conditions for minimisers of the functional to satisfy a corresponding variational inequality.

Theorem 2.5 ([32, Theorem 46.A(a)]) *Let $F\colon C \to \mathbb{R}$ be a functional on the convex non-empty set C of a real locally convex space X, and let $b \in X'$ be a given element. Suppose the Gateaux-derivative F' exists on C. Then any solution f of*

$$\min_{f \in C} \left\{ F(f) - \langle b, f \rangle_{X' \times X} \right\} \tag{2.40}$$

satisfies the variational inequality

$$\langle F'(f) - b, h - f \rangle_{X' \times X} \geq 0, \qquad \text{for all } h \in C. \tag{2.41}$$

3 A novel representation of the assignment flow

Let $J\colon \mathcal{W} \to \mathbb{R}$ be a smooth function on the assignment manifold (2.21) and denote the Riemannian gradient of J at $W \in \mathcal{W}$ induced by the Fisher–Rao metric (2.15) by $\operatorname{grad} J(W) \in \mathcal{T}_0$. In view of the embedding (2.22), we can also compute the Euclidean gradient of J denoted by $\partial J(W) \in \mathbb{R}^{n \times c}$. These two gradients are related by [6, Proposition 1]

$$\operatorname{grad} J(W) = R_W \, \partial J(W), \qquad W \in \mathcal{W}, \tag{3.1}$$

where $R_W\colon \mathbb{R}^{n \times c} \to \mathcal{T}_0$ is the product map obtained by applying R_{W_i} from (2.16) to every row vector indexed by $i \in \mathcal{I}$. This relation raises the natural question: Is there a potential J such that the assignment flow (2.32) is a Riemannian gradient descent flow with respect to J, that is, does $R_W S(W) = -\operatorname{grad} J(W)$ hold?

We next show that such a potential does not exist in general (Section 3.1). However, in Section 3.2, we derive a novel representation by decoupling the assignment flow into two separate flows, where one flow steers the other and in this sense dominates the assignment flow. Under the additional assumption that the weights ω_{ij} of the similarity map $S(W)$ in (2.29) are symmetric, we show that the dominating flow is a Riemannian gradient flow induced by a potential. This result is the basis for the continuous-domain formulation of the assignment flow studied in the subsequent sections.

3.1 Non-potential flow

We next show (Theorem 3.5) that under some mild assumptions on $D_{\mathcal{F}}$ (2.26) which are always fulfilled in practice, no potential J exists that induces the assignment flow. In order to prove this result, we first derive some properties of the mapping exp given by (2.19c) as well as explicit expressions of the differential $dS(W)$ of the similarity map (2.29) and its transpose $dS(W)^\top$ with respect to the standard Euclidean structure on $\mathbb{R}^{n \times c}$.

Lemma 3.1 *The following properties hold for* $\exp_p$ *and its inverse* (2.19c), (2.19d).

(1) *For every $p \in \mathcal{S}$, the map* $\exp_p\colon \mathbb{R}^c \to \mathcal{S}$ *can be expressed by*

$$v \mapsto \exp_p(v) = \frac{p e^v}{\langle p, e^v \rangle}. \tag{3.2}$$

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

Its restriction to T_0, $\exp_p\colon T_0 \to \mathcal{S}$, *is a diffeomorphism. The differential of* $\exp_p$ *and* $\exp_p^{-1}$ *at* $v \in T_0$ *and* $q \in \mathcal{S}$, *respectively, are given by*

$$d\exp_p(v)[u] = R_{\exp_p(v)}[u] \quad \text{and} \quad d\exp_p^{-1}(q)[u] = \Pi_0 \begin{bmatrix} u \\ q \end{bmatrix} \quad \text{for all } u \in T_0. \tag{3.3}$$

(2) *Let* $p, q \in \mathcal{S}$. *Then,* $\operatorname{Exp}_p^{-1}(q) = R_p \exp_p^{-1}(q)$.

(3) *Let* $q \in \mathcal{S}$. *If the linear map* R_q *from* (2.16) *is restricted to* T_0, *then* $R_q\colon T_0 \to T_0$ *is a linear isomorphism with inverse given by* $(R_q|_{T_0})^{-1}(u) = \Pi_0[\frac{u}{q}]$ *for all* $u \in T_0$.

(4) *If* $\mathbb{R}^c$ *is viewed as an abelian group, then* $\exp\colon \mathbb{R}^c \times \mathcal{S} \to \mathcal{S}$ *given by* $(v, p) \mapsto \exp_p(v)$ *defines a Lie-group action, that is,*

$$\exp_p(v+u) = \exp_{\exp_p(u)}(v) \quad \text{and} \quad \exp_p(0) = p \qquad \text{for all } v, u \in T_0 \text{ and } p \in \mathcal{S}. \tag{3.4}$$

Furthermore, the following identities follow for all $p, q, a \in \mathcal{S}$ *and* $v \in \mathbb{R}^c$

$$\exp_p(v) = \exp_q\left(v + \exp_q^{-1}(p)\right) \tag{3.5a}$$

$$\exp_q^{-1}(p) = -\exp_p^{-1}(q) \tag{3.5b}$$

$$\exp_q^{-1}(a) = \exp_p^{-1}(a) - \exp_p^{-1}(q). \tag{3.5c}$$

Proof (1): We have $\operatorname{Exp}_p(v + \lambda p) = \operatorname{Exp}_p(v)$ for every $p \in \mathcal{S}$, $v \in T_0$ and $\lambda \in \mathbb{R}$, as a simple computation using definition (2.19a) of Exp_p directly shows. Therefore, for every $v \in T_0$,

$$\exp_p(v) = \operatorname{Exp}_p(R_p v) = \operatorname{Exp}_p(pv - \langle v, p\rangle p) = \operatorname{Exp}_p(pv) = \frac{pe^v}{\langle p, e^v\rangle}. \tag{3.6}$$

If we restrict $\exp_p$ to T_0, then an inverse is explicitly given by (2.19c). The differentials (3.3) result from a standard computation.

(2): The formula is a direct consequence of the formulas for $\operatorname{Exp}_p^{-1}$ and $\exp_p^{-1}$ given in (2.19b) and (2.19c), together with the fact (2.17).

(3): Fix any $p \in \mathcal{S}$ and set $v_q := \exp_p^{-1}(q)$ for $q \in \mathcal{S}$. Since $\exp_p\colon T_0 \to \mathcal{S}$ is a diffeomorphism, the differential $d\exp_p(v_q)\colon T_0 \to T_0$ is an isomorphism. By (3.3), we have $R_q[u] = R_{\exp_p(v_q)}[u] = d\exp_p(v_q)[u]$ for all $u \in T_0$, showing that R_q is an isomorphism with the corresponding inverse.

(4): Properties (3.4) defining the group action are directly verified using (3.2). Now, suppose $p, q, a \in \mathcal{S}$ and $v \in \mathbb{R}^c$ are arbitrary. Since $\exp_q\colon T_0 \to \mathcal{S}$ is a diffeomorphism, we have $p = \exp_q\left(\exp_q^{-1}(p)\right)$ and by the group action property

$$\exp_p(v) = \exp_{\exp_q\left(\exp_q^{-1}(p)\right)}(v) = \exp_q\left(v + \exp_q^{-1}(p)\right), \tag{3.7}$$

which proves (3.5a). To show (3.5b), set $v_a := \exp_p^{-1}(a)$ and substitute this vector into (3.5a). Applying $\exp_q^{-1}$ to both sides then gives

$$\exp_q^{-1}(a) = \exp_q^{-1}\left(\exp_p(v_a)\right) = v_a + \exp_q^{-1}(p) = \exp_p^{-1}(a) + \exp_q^{-1}(p). \tag{3.8}$$

Setting $a = q$ in this equation, we obtain (3.5b) from

$$0 = \exp_q^{-1}(q) = \exp_p^{-1}(q) + \exp_q^{-1}(p). \tag{3.9}$$

Using $\exp_q^{-1}(p) = -\exp_p^{-1}(q)$ in (3.8) yields (3.5c). □

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

Lemma 3.2 *The ith component of the similarity map $S(W)$ defined by (2.29) can equivalently be expressed as*

$$S_i(W) = \exp_{\mathbb{1}_{\mathcal{S}}}\left(\sum_{j\in\mathcal{N}_i}\omega_{ij}\left(\exp_{\mathbb{1}_{\mathcal{S}}}^{-1}(W_j) - \frac{1}{\rho}D_{\mathcal{F};j}\right)\right) \quad \text{for all } i\in\mathcal{I} \text{ and } W\in\mathcal{W}. \tag{3.10}$$

Proof Consider the expression $\operatorname{Exp}_{W_i}^{-1}\big(L_j(W_j)\big)$ in the sum of the definition (2.29) of $S_i(W)$. Using (3.2) and (3.5a), the likelihood (2.27) can be expressed as

$$L_j(W_j) = \exp_{W_j}\left(-\frac{1}{\rho}D_{\mathcal{F};j}\right) = \exp_{W_i}\left(\exp_{W_i}^{-1}(W_j) - \frac{1}{\rho}D_{\mathcal{F};j}\right). \tag{3.11}$$

In the following, we set

$$V_k = \exp_{\mathbb{1}_{\mathcal{S}}}^{-1}(W_k) \qquad \text{for all } k\in\mathcal{I}. \tag{3.12}$$

With this and (3.5c), we have

$$\exp_{W_i}^{-1}(W_j) = \exp_{\mathbb{1}_{\mathcal{S}}}^{-1}(W_j) - \exp_{\mathbb{1}_{\mathcal{S}}}^{-1}(W_i) = V_j - V_i. \tag{3.13}$$

The two previous identities and Lemma 3.1(2) give

$$\operatorname{Exp}_{W_i}^{-1}\big(L_j(W_j)\big) = R_{W_i}\Big[\exp_{W_i}^{-1}\big(L_j(W_j)\big)\Big] = R_{W_i}\left[\exp_{W_i}^{-1}(W_j) - \frac{1}{\rho}D_{\mathcal{F};j}\right] \tag{3.14a}$$

$$= R_{W_i}\left[V_j - V_i - \frac{1}{\rho}D_{\mathcal{F};j}\right]. \tag{3.14b}$$

The sum over the neighbouring nodes $\mathcal{N}_i$ in the definition (2.29) of $S_i(W)$ can therefore be rewritten as

$$\sum_{j\in\mathcal{N}_i}\omega_{ij}\operatorname{Exp}_{W_i}^{-1}\big(L_j(W_j)\big) = \sum_{j\in\mathcal{N}_i}\omega_{ij}R_{W_i}\left[V_j - V_i - \frac{1}{\rho}D_{\mathcal{F};j}\right] \tag{3.15a}$$

$$= R_{W_i}\left[-V_i + \sum_{j\in\mathcal{N}_i}\omega_{ij}\left(V_j - \frac{1}{\rho}D_{\mathcal{F};j}\right)\right], \tag{3.15b}$$

where we used $\sum_{j\in\mathcal{N}_i}\omega_{ij} = 1$ for the last equation. Setting $Y_i := \sum_{j\in\mathcal{N}_i}\omega_{ij}\Big(V_j - \frac{1}{\rho}D_{\mathcal{F};j}\Big)$, we then have

$$S_i(W) = \operatorname{Exp}_{W_i}\big(R_{W_i}\big[-V_i + Y_i\big]\big) = \exp_{W_i}\big(-V_i + Y_i\big) = \exp_{\mathbb{1}_{\mathcal{S}}}\big(Y_i\big), \tag{3.16}$$

where the last equality again follows from (3.5a) together with the definition (3.12) of V_i. □

Lemma 3.3 *The ith component of the differential of the similarity map $S(W)\in\mathcal{W}$ is given by*

$$dS_i(W)[X] = \sum_{j\in\mathcal{N}_i}\omega_{ij}R_{S_i(W)}\left[\frac{X_j}{W_j}\right] \quad \text{for all } X\in\mathcal{T}_0 \text{ and } i\in\mathcal{I}. \tag{3.17}$$

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

Furthermore, the ith component of the adjoint differential $dS(W)^\top : \mathcal{T}_0 \to \mathcal{T}_0$ *with respect to the standard Euclidean inner product on* $\mathcal{T}_0 \subset \mathbb{R}^{n\times c}$ *is given by*

$$dS_i(W)^\top[X] = \sum_{j\in\mathcal{N}_i} \omega_{ji} \Pi_0 \left[\frac{R_{S_j(W)} X_j}{W_i}\right] \quad \text{for every } X \in \mathcal{T}_0 \text{ and } i \in \mathcal{I}. \tag{3.18}$$

Proof Define the map $F_i : \mathcal{W} \to \mathbb{R}^c$ by $F_i(W) := \sum_{j\in\mathcal{N}_i} \omega_{ij} \Big(\exp_{\mathbb{1}_\mathcal{S}}^{-1}(W_j) - \frac{1}{\rho} D_{\mathcal{F};j} \Big) \in \mathbb{R}^c$ for all $W \in \mathcal{W}$. Let $\gamma : (-\varepsilon, \varepsilon) \to \mathcal{W}$ be a smooth curve, with $\varepsilon > 0$, $\gamma(0) = W$ and $\dot\gamma(0) = X$. By (3.3), we then have

$$dF_i(W)[X] = \frac{d}{dt} F_i(\gamma(t))\big|_{t=0} = \sum_{j\in\mathcal{N}_i} \omega_{ij} \frac{d}{dt} \exp_{\mathbb{1}_\mathcal{S}}^{-1}(\gamma_j(t))\big|_{t=0} = \sum_{j\in\mathcal{N}_i} \omega_{ij} \Pi_0 \left[\frac{X_j}{W_j}\right]. \tag{3.19}$$

Due to Lemma 3.2, we can express the ith row of the similarity map as $S_i(W) = \exp_{\mathbb{1}_\mathcal{S}}\big(F_i(W)\big)$. Therefore, the differential of S_i is given by

$$dS_i(W)[X] = d\exp_{\mathbb{1}_\mathcal{S}}(F_i(W))\big[dF_i(W)[X]\big] = R_{\exp_{\mathbb{1}_\mathcal{S}}(F_i(W))}\big[dF_i(W)[X]\big] \tag{3.20a}$$

$$= R_{S_i(W)}\left[\sum_{j\in\mathcal{N}_i} \omega_{ij} \Pi_0 \left[\frac{X_j}{W_j}\right]\right] = \sum_{j\in\mathcal{N}_i} \omega_{ij} R_{S_i(W)} \left[\frac{X_j}{W_j}\right], \tag{3.20b}$$

where we used $R_{S_i(W)} \Pi_0 = R_{S_i(W)}$ from (2.17) to obtain the last equation.

Now let $W \in \mathcal{W}$ and $X, Y \in \mathcal{T}_0$ be arbitrary. By assumption on the neighbourhood structure (2.7), we have $j \in \mathcal{N}_i$ if and only if $i \in \mathcal{N}_j$, that is, $\psi_{\mathcal{N}_i}(j) = \psi_{\mathcal{N}_j}(i)$. Since $R_{S_i(W)} \in \mathbb{R}^{c\times c}$ is a symmetric matrix, we obtain

$$\langle dS(W)[X], Y\rangle = \sum_{i\in\mathcal{I}} \langle dS_i(W)[X], Y_i\rangle = \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{N}_i} \omega_{ij} \left\langle R_{S_i(W)} \left[\frac{X_j}{W_j}\right], Y_i \right\rangle \tag{3.21a}$$

$$= \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{I}} \psi_{\mathcal{N}_i}(j) \omega_{ij} \left\langle \frac{X_j}{W_j}, R_{S_i(W)}[Y_i] \right\rangle = \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{I}} \psi_{\mathcal{N}_j}(i) \omega_{ij} \left\langle X_j, \frac{R_{S_i(W)}[Y_i]}{W_j} \right\rangle \tag{3.21b}$$

$$= \sum_{j\in\mathcal{I}} \sum_{i\in\mathcal{N}_j} \omega_{ij} \left\langle X_j, \Pi_0 \left[\frac{R_{S_i(W)}[Y_i]}{W_j}\right] \right\rangle = \sum_{j\in\mathcal{I}} \left\langle X_j, \sum_{i\in\mathcal{N}_j} \omega_{ij} \Pi_0 \left[\frac{R_{S_i(W)}[Y_i]}{W_j}\right] \right\rangle. \tag{3.21c}$$

On the other hand, we have

$$\langle dS(W)[X], Y\rangle = \big\langle X, dS(W)^\top[Y] \big\rangle = \sum_{j\in\mathcal{I}} \big\langle X_j, dS_j(W)^\top[Y] \big\rangle. \tag{3.22}$$

Because (3.21) and (3.22) hold for all $X, Y \in \mathcal{T}_0$, the formula for $dS_i(W)^\top[X]$ is proven. □

In order to prepare the main result of this section, the non-existence of a potential, we first prove a technical statement based on the assumption that all neighbourhoods contain at least three vertices, that is,

$$|\mathcal{N}_i| \geq 3, \quad \text{for all } i \in \mathcal{I}. \tag{3.23}$$

In almost all practical situations, this will be the case.

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

Lemma 3.4 *Suppose assumption* (3.23) *holds and let* $i, j \in \mathcal{I}$ *be two vertices with* $j \in \mathcal{N}_i$. *Then, for any* $p \in \mathcal{S}$, *there is a point* $W \in \mathcal{W}$ *such that the equations*

$$S_j(W) = W_i, \qquad S_i(W) = \mathbb{1}_{\mathcal{S}}, \qquad W_j = p \tag{3.24}$$

hold.

Proof Due to the additional assumption (3.23), there is an $r \in \mathcal{N}_i$ with $r \neq i, j$. For all $k \in \mathcal{I}$ with $k \neq r, i, j$, the rows $W_k \in \mathcal{S}$ can be chosen arbitrarily, for example, $W_k = \mathbb{1}_{\mathcal{S}}$. As $W_j = p$ has to hold, we can only vary W_i and W_r in order to satisfy the remaining two conditions. To this end, parametrise

$$W_i = \exp_{\mathbb{1}_{\mathcal{S}}}(a) \quad \text{and} \quad W_r = \exp_{\mathbb{1}_{\mathcal{S}}}(b) \quad \text{with} \quad a, b \in T_0. \tag{3.25}$$

In the following, two linear equations are derived for determining the values for a and b. Due to $r \in \mathcal{N}_i$ and Lemma 3.2, applying $\exp_{\mathbb{1}_{\mathcal{S}}}^{-1}$ to $S_i(W)$ yields

$$\exp_{\mathbb{1}_{\mathcal{S}}}^{-1}\big(S_i(W)\big) = \sum_{k \in \mathcal{N}_i} \omega_{ik} \Big(\exp_{\mathbb{1}_{\mathcal{S}}}^{-1}(W_k) - \frac{1}{\rho} D_{\mathcal{F};k} \Big) = \omega_{ii} a + \omega_{ir} b + \alpha, \tag{3.26}$$

where $\alpha \in T_0$ collects all terms not involving a or b. Thus, the second condition (3.24) is equivalent to

$$0 = \exp_{\mathbb{1}_{\mathcal{S}}}^{-1}\big(\mathbb{1}_{\mathcal{S}}\big) = \omega_{ii} a + \omega_{ir} b + \alpha. \tag{3.27}$$

Regarding the transformation of the first condition (3.24), note that although $r, j \in \mathcal{N}_i$, the inclusion $r \in \mathcal{N}_j$ does *not* necessarily hold. Thus, applying $\exp_{\mathbb{1}_{\mathcal{S}}}^{-1}$ to $S_j(W)$ leads to

$$\exp_{\mathbb{1}_{\mathcal{S}}}^{-1}\big(S_j(W)\big) = \sum_{k \in \mathcal{N}_j} \omega_{jk} \Big(\exp_{\mathbb{1}_{\mathcal{S}}}^{-1}(W_k) - \frac{1}{\rho} D_{\mathcal{F};k} \Big) = \omega_{ji} a + \psi_{\mathcal{N}_j}(r) \omega_{jr} b + \beta, \tag{3.28}$$

where $\beta \in T_0$ again collects all terms not involving a or b, and $\psi_{\mathcal{N}_j}$ is the indicator function defined in Section 2.1. Replacing $S_j(W)$ by W_i on the left-hand side of (3.28) due to the first condition (3.24), and using the first relation (3.25), we obtain

$$a = \exp_{\mathbb{1}_{\mathcal{S}}}^{-1}\big(W_i\big) = \exp_{\mathbb{1}_{\mathcal{S}}}^{-1}\big(S_j(W)\big) = \omega_{ji} a + \psi_{\mathcal{N}_j}(r) \omega_{jr} b + \beta. \tag{3.29}$$

Using (3.26) to solve for a and substituting the result into (3.29) to solve for b yield the unique solution

$$b = \big((1 - \omega_{ji})\omega_{ir} + \psi_{\mathcal{N}_j}(r)\omega_{jr}\omega_{ii}\big)^{-1}\big(\omega_{ii}\beta + (1 - \omega_{ji})\alpha\big), \tag{3.30a}$$

$$a = -\omega_{ii}^{-1}\big(\omega_{ir} b + \alpha\big), \tag{3.30b}$$

which proves the existence of a $W \in \mathcal{W}$ satisfying (3.24). □

We are now prepared to prove non-existence of a potential for the assignment flow. This will be done under the additional assumption that for at least one node i the corresponding distance vector $D_{\mathcal{F};i}$ (2.25) is not constant, that is, there is at least one preferred label choice based just on the given data and prototypes, at some vertex. As any measured data in reality contain measurement noise, this assumption is almost always fulfilled in practice.

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

Theorem 3.5 (Non-existence of a potential) *Suppose assumption* (3.23) *holds and there exists a node* $i \in \mathcal{I}$ *such that the distance vector* $D_{\mathcal{F};i}$ *is not constant, that is,* $D_{\mathcal{F};i} \notin \mathbb{R}\mathbb{1}$. *Then no potential* $J\colon \mathcal{W} \to \mathbb{R}$ *exists satisfying*

$$R_W S(W) = -\operatorname{grad} J(W), \tag{3.31}$$

that is, the assignment flow (2.32) *is not a Riemannian gradient descent flow with respect to the Fisher–Rao metric*

Proof To simplify notation during the proof, we write D_j instead of $D_{\mathcal{F};j}$ for the jth row of the distance matrix. To prove the statement, we show the contrapositive: if a potential J exists for the assignment flow, then for every $j \in \mathcal{I}$ the distance vector D_j needs to be constant, that is, $D_j \in \mathbb{R}\mathbb{1}_c$. Therefore, suppose a potential J exists and let $j \in \mathcal{I}$ be arbitrary. The proof of $D_j \in \mathbb{R}\mathbb{1}_c$ is subdivided into two steps. First, the necessary condition (3.41) at vertex j is derived using Lemma 3.4. Second, based on this condition, $D_j \in \mathbb{R}\mathbb{1}$ is shown via a limiting process.

Step 1: By (2.17), we have $R_W S(W) = R_W \Pi_0 S(W)$ and $R_W\colon \mathcal{T}_0 \to \mathcal{T}_0$ is a linear isomorphism (Lemma 3.1(3)). Therefore, we can get rid of the replicator operator in (3.1) by applying $(R_W|_{\mathcal{T}_0})^{-1}$ to both sides of the equation, that is,

$$R_W S(W) = -\operatorname{grad} J(W) \overset{(3.1)}{=} -R_W\, \partial J(W) \quad \Leftrightarrow \quad \Pi_0 S(W) = -\Pi_0\, \partial J(W) \in \mathcal{T}_0. \tag{3.32}$$

The negative Euclidean Hessian of J is then given by

$$-\Pi_0 \operatorname{Hess} J(W) = d\big(-\Pi_0\, \partial J\big)(W) \overset{(3.32)}{=} d(\Pi_0 \circ S)(W) = \Pi_0 dS(W) = dS(W), \tag{3.33}$$

where the last equality follows from $dS(W)\colon \mathcal{T}_0 \to \mathcal{T}_0$. Furthermore, $\operatorname{Hess} J(W)$ and therefore also $dS(W)$ must be symmetric with respect to the Euclidean inner product on $\mathcal{T}_0$. Thus, under the assumption that a potential J exists, $dS(W)[X] - dS(W)^\top[X] = 0$ holds, or equivalently

$$dS_i(W)[X] - dS_i(W)^\top[X] = 0 \quad \text{for every} \quad i \in \mathcal{I}. \tag{3.34}$$

For the following, choose any $i \in \mathcal{N}_j$ and fix it. By (2.8), we have $j \in \mathcal{N}_i$. Next, let $p \in \mathcal{S}$ be arbitrary. According to Lemma 3.4, there exists a $W^{(p)} \in \mathcal{W}$ with

$$S_j(W^{(p)}) = W_i^{(p)}, \quad S_i(W^{(p)}) = \mathbb{1}_{\mathcal{S}} \quad \text{and} \quad W_j^{(p)} = p. \tag{3.35}$$

For $l, s \in [c]$, set

$$u := e_s - e_l \in T_0, \tag{3.36}$$

with $e_s, e_l \in \mathcal{B}_c$ and define the tangent vector

$$X^u \in \mathcal{T}_0 \quad \text{with} \quad X^u = \begin{cases} u, & \text{if } k = j \\ 0, & \text{if } k \neq j. \end{cases} \quad \text{for all } k \in \mathcal{I}. \tag{3.37}$$

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

Due to the expressions for $dS_i\left(W^{(p)}\right)$ and $dS_i\left(W^{(p)}\right)^\top$ from Lemma 3.3 and as a consequence of the choice of X^u, we obtain

$$dS_i\left(W^{(p)}\right)[X^u] - dS_i\left(W^{(p)}\right)^\top[X^u] = \omega_{ij} R_{S_i(W^{(p)})}\left[\frac{u}{W_j^{(p)}}\right] - \omega_{ji}\Pi_0\left[\frac{R_{S_j(W^{(p)})}u}{W_i^{(p)}}\right] \tag{3.38a}$$

$$\overset{(3.35)}{=} \omega_{ij} R_{\mathbb{1}_{\mathcal{S}}}\left[\frac{u}{p}\right] - \omega_{ji}\Pi_0\left[u - \langle u, S_j(W^{(p)})\rangle \mathbb{1}_c\right], \tag{3.38b}$$

and because of $\Pi_0[\mathbb{1}_c] = 0$ together with the identity $R_{\mathbb{1}_{\mathcal{S}}} = c^{-1}\Pi_0$ from (2.17), we further get

$$= \Pi_0\left[\left(c^{-1}\omega_{ij}p^{-1} - \omega_{ji}\mathbb{1}_c\right)u\right] \overset{(3.34)}{=} 0. \tag{3.38c}$$

Temporarily setting

$$z := c^{-1}\omega_{ij}p^{-1} - \omega_{ji}\mathbb{1}_c, \tag{3.39}$$

Equation (3.38c) reads $0 = \Pi_0[zu]$. As a consequence of the fact that Π_0 is the orthogonal projection onto T_0 and $u^2 = u\,u = e_l + e_s$ (componentwise multiplication) by the definition of u in (3.36), it follows that

$$0 = \langle \Pi_0[zu], u\rangle = \langle uz, u\rangle = \langle z, u^2\rangle = z_l + z_s. \tag{3.40}$$

Substituting (3.39) back into (3.40) finally gives the desired condition

$$0 = c^{-1}\omega_{ij}\left(p_l^{-1} - p_s^{-1}\right) - 2\omega_{ji} \quad \text{for all } l, s \in [c],\ p \in \mathcal{S},\ i \in \mathcal{N}_j. \tag{3.41}$$

Step 2: Let $\tau > 0$ be arbitrary and set

$$p := \exp_{\mathbb{1}_{\mathcal{S}}}\left(\frac{1}{\tau}D_j\right) = \frac{e^{\frac{1}{\tau}D_j}}{\left\langle e^{\frac{1}{\tau}D_j}, \mathbb{1}_c\right\rangle}. \tag{3.42}$$

Using this definition of p in (3.41) and rearranging the resulting equation yield

$$1 = (2c\omega_{ji})^{-1}\omega_{ij}\left\langle e^{\frac{1}{\tau}D_j}, \mathbb{1}_c\right\rangle\left(e^{-\frac{1}{\tau}D_{jl}} + e^{-\frac{1}{\tau}D_{js}}\right) \tag{3.43}$$

for all $l, s \in [c]$. Applying the log on both sides and multiplying by τ gives

$$0 = \tau\log\left((2c\omega_{ji})^{-1}\omega_{ij}\right) + \tau\log\left\langle e^{\frac{1}{\tau}D_j}, \mathbb{1}_c\right\rangle + \tau\log\left(e^{-\frac{1}{\tau}D_{jl}} + e^{-\frac{1}{\tau}D_{js}}\right). \tag{3.44}$$

for all $\tau > 0$. By [25, Example 1.30],

$$\lim_{\tau\to 0}\tau\log\left\langle e^{\frac{1}{\tau}x}, \mathbb{1}_c\right\rangle = \max_{k\in[d]} x_k \tag{3.45}$$

holds for any $x \in \mathbb{R}^d$. Therefore, taking the limit $\tau \to 0$ in (3.44) results in the condition

$$0 = \max_{k\in[c]} D_{jk} + \max\{-D_{jl}, -D_{js}\} = \max_{k\in[c]} D_{jk} - \min\left\{D_{jl}, D_{js}\right\}. \tag{3.46}$$

To finish the proof, let $l \in \operatorname{argmax}_{k\in[c]} D_{jk}$ and $s \in [c]$ be arbitrary. Then,

$$D_{jl} = \max_{k\in[c]} D_{jk} = \min\{D_{jl}, D_{js}\} = D_{js}, \quad \text{for all } s \in [c], \tag{3.47}$$

proving that $D_j \in \mathbb{R}\mathbb{1}_c$. □

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

3.2 S-parametrisation

Even though Theorem 3.5 says that no potential exists for the assignment flow in general, we reveal in this section a 'hidden' potential flow under an additional assumption. To this end, we decouple the assignment flow into two components and show that one component depends on the second one. The dominating second one, therefore, provides a new parametrisation of the assignment flow. Assuming symmetry of the averaging matrix defined below by (3.48), the dominating flow becomes a Riemannian gradient descent flow. The corresponding potential defined on a continuous domain will be studied in subsequent sections.

For notational efficiency, we collect all weights (2.28) into the *averaging matrix*

$$\Omega^{\omega} \in \mathbb{R}^{n\times n} \text{ with } \Omega^{\omega}_{ij} := \psi_{\mathcal{N}_i}(j)\omega_{ij} = \begin{cases} \omega_{ij} & \text{if } j \in \mathcal{N}_i, \\ 0 & \text{else} \end{cases}, \quad \text{for } i,j \in \mathcal{I}. \tag{3.48}$$

Ω^{ω} encodes the spatial structure of the graph and the weights. For an arbitrary matrix $M \in \mathbb{R}^{n\times c}$, the average of its row vectors using the weights indexed by the neighbourhood $\mathcal{N}_i$ is given by

$$\sum_{k\in\mathcal{N}_i} \omega_{ik} M_k = \sum_{k\in\mathcal{I}} \Omega^{\omega}_{ik} M_k = (\Omega^{\omega} M)_i. \tag{3.49}$$

Thus, all row vector averages are given as row vectors of the matrix $\Omega^{\omega} M \in \mathbb{R}^{n\times c}$.

We now introduce a new representation of the assignment flow.

Proposition 3.6 *The assignment flow* (2.32) *is equivalent to the system*

$$\dot{W} = R_W[\overline{S}] \qquad \textit{with} \quad W(0) = \mathbb{1}_{\mathcal{W}} \tag{3.50a}$$

$$\dot{\overline{S}} = R_{\overline{S}}[\Omega^{\omega}\overline{S}] \qquad \textit{with} \quad \overline{S}(0) = S(\mathbb{1}_{\mathcal{W}}). \tag{3.50b}$$

Remark 3.7 *We observe that the flow $W(t)$ is completely determined by $\overline{S}(t)$. In the following, we refer to the dominating part* (3.50b) *as the S-flow.*

Proof Let $W(t)$ be a solution of the assignment flow, that is, $\dot{W}_i = R_{W_i} S_i(W)$ for all $i \in \mathcal{I}$. Set $\overline{S}(t) := S(W(t))$. Then, (3.50a) is immediate from the assumption on W. Using the expression for $dS_i(W)$ from Lemma 3.3 gives

$$\dot{\overline{S}}_i = \frac{d}{dt} S(W)_i = dS_i(W)[\dot{W}] = \sum_{j\in\mathcal{N}_i} \omega_{ij} R_{S_i(W)} \left[\frac{\dot{W}_j}{W_j}\right]. \tag{3.51}$$

Since W solves the assignment flow and $R_{S_i(W)} = R_{S_i(W)}\Pi_0$ by (2.17) with $\ker(\Pi_0) = \mathbb{R}\mathbb{1}_c$, it follows using the explicit expresssion (2.16) of $R_{S_i(W)}$ that

$$R_{S_i(W)} \left[\frac{\dot{W}_j}{W_j}\right] = R_{S_i(W)} \left[\frac{R_{W_j} S_j(W)}{W_j}\right] = R_{S_i(W)} \left[S_j(W) - \langle W_j, S_j(W)\rangle \mathbb{1}_c\right] \tag{3.52a}$$

$$= R_{S_i(W)}\big[S_j(W)\big]. \tag{3.52b}$$

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

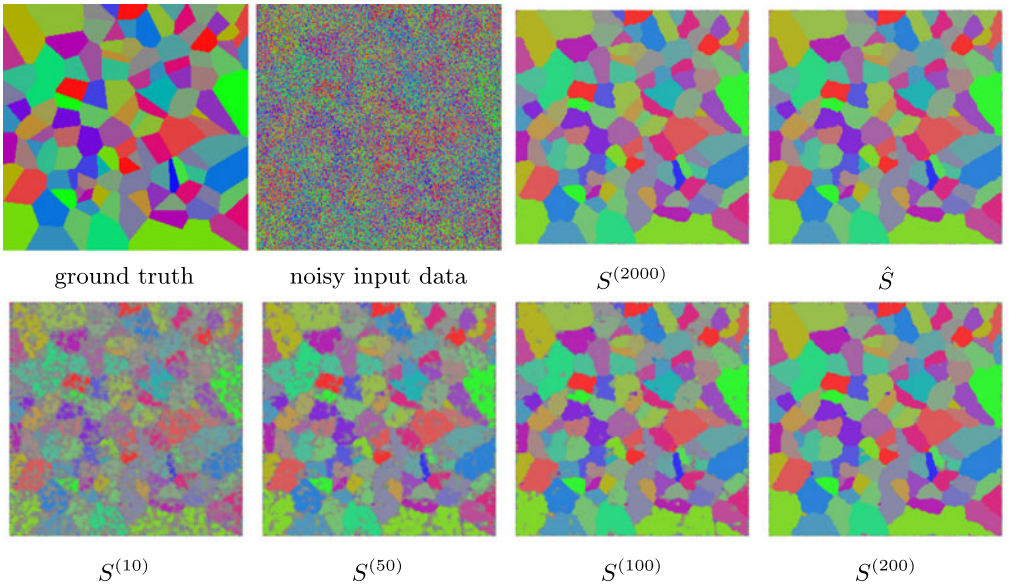

FIGURE 2. Illustration of the S-flow (3.50b) using the geometric Euler scheme (3.55) for numerical integration. Parameter values: $\rho = 0.1$ and $h_k = h = 0.1$, for every iteration k. Top, from left to right: Ground truth, noisy input data, iterate $S^{(2000)}$ and the corresponding integral labelling $\hat{S}$ resulting from $S^{(2000)}$ by rounding to the nearest integral label at each pixel, that is, choosing $\operatorname{argmax}_{r\in[c]} S_{ir}^{(2000)}$, for every $i \in \mathcal{I}$. The remaining minimal noise at the boundary of $S^{(2000)}$ is a result from the higher centre weight ω_{ii} due to the symmetrisation of the uniform weights (3.54). Bottom, from left to right: The iterates $S^{(10)}$, $S^{(50)}$, $S^{(100)}$ and $S^{(200)}$. The sequence $S^{(k)}$ converges fairly quickly in the first 200 iterations to a reasonable partition. However, 1800 additional iterations are needed to eliminate most of the still visible very small blob-like assignments. On the other hand, each iteration only requires local communication and hence can be efficiently carried out using graphics processing units (GPUs).

Back-substitution of this identity into (3.51), pulling the linear map $R_{S_i(W)}$ out of the sum and keeping $S_i(W) = \overline{S}_i$ in mind result in

$$\dot{\overline{S}}_i = R_{\overline{S}_i}\left[\sum_{j\in\mathcal{N}_i} \omega_{ij}\overline{S}_j\right] = R_{\overline{S}_i}[(\Omega^{\omega}\overline{S})_i] \quad \text{for all } i \in \mathcal{I}. \tag{3.53}$$

Collecting these vectors as row vectors of the matrix $\dot{\overline{S}}$ gives (3.50b). □

Remark 3.8 *Henceforth, we write S for the S-flow $\overline{S}$ to stress the underlying connection to the assignment flow and to simplify the notation.*

The basic behaviour of the S-flow is illustrated by the following numerical experiment, shown as Figure 2. There are 31 labels visualised by 31 colours. The colour of each pixel of 'ground truth' represents a vertex (unit vector) of the corresponding probability simplex. The ℓ_1-norm was used to compute the distance matrix. We adopted this experiment from [6, Figure 6] because the uniform distances between labels enable to assess the regularisation properties of the assignment flow in an unbiased way.

The spatial 256×256 pixel grid is represented by a grid graph with vertex set $\mathcal{I} = [256] \times [256]$. The neighbourhood $\mathcal{N}_i$ at pixel i is given by all pixels $j \in \mathcal{I}$ contained in the 5×5 pixel

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

patch centred at i. For pixels i sufficiently far inside the image, this results in $|\mathcal{N}_i| = 5 \times 5$, while $|\mathcal{N}_i| < 5 \times 5$ for pixels close to the boundary. For example, $|\mathcal{N}_i| = 9 < 25$ at the upper left corner pixel i. Due to these boundary effects in terms of neighbourhood sizes, *symmetric uniform weights* defined by

$$\omega_{ij} := \begin{cases} \dfrac{1}{5^2}, & \text{for } i \neq j \\ \dfrac{1}{5^2}\big(1 + 5^2 - |\mathcal{N}_i|\big), & \text{for } i = j \end{cases} \tag{3.54}$$

were used. This results in uniform weights $\omega_{ij} = \frac{1}{5^2}$ for $i \in \mathcal{I}$ in the interior and suitable adjustments at pixels i at or close to the boundary, while preserving the overall symmetry of the weights, that is, $\omega_{ij} = \omega_{ji}$, which will become relevant below. Due to this symmetrisation, boundary pixels have a higher centre weight ω_{ii}, e.g. $\omega_{ii} = 17/5^2 = 0.68$ for the upper-left corner pixel i.

For numerically integrating the S-flow (3.50b) on $\mathcal{W}$, the basic geometric Euler method from [35] was used. The corresponding numerical update scheme is given by

$$S^{(k+1)} = \exp_{S^{(k)}}\left(h_k \Omega^\omega S^{(k)}\right) \qquad S^{(0)} := S(0) = S(\mathbb{1}_{\mathcal{W}}), \tag{3.55}$$

with step size $h_k > 0$. All relevant parameter values are specified in the caption of Figure 2.

We next show that the S-flow which essentially determines the assignment flow (Remark 3.8) becomes a Riemannian descent flow under the additional assumption that the averaging matrix (3.48) is symmetric.

Proposition 3.9 *Suppose the weights defining the similarity map in* (2.29) *are symmetric, that is,* $(\Omega^\omega)^\top = \Omega^\omega$. *Then the S-flow* (3.50b) *is a Riemannian gradient descent flow* $\dot S = -\operatorname{grad} J(S)$, *induced by the potential*

$$J(S) := -\frac{1}{2}\langle S, \Omega^\omega S\rangle, \quad S \in \mathcal{W}. \tag{3.56}$$

Proof Let $\gamma : (-\varepsilon, \varepsilon) \to \mathcal{W}$, $\varepsilon > 0$, be any smooth curve with $\dot\gamma(0) = V \in \mathbb{R}^{n\times c}$ and $\gamma(0) = S$. By the symmetry of Ω^ω, we have $\langle \partial J(S), V\rangle = dJ(S)[V] = \frac{d}{dt} J(\gamma(t))\big|_{t=0} = -\langle \Omega^\omega S, V\rangle$ for all $V \in \mathbb{R}^{n\times c}$. Therefore, $\partial J(S) = -\Omega^\omega S$. Thus, the Riemannian gradient is given by $\operatorname{grad} J(S) = R_S[\partial J(S)] = -R_S[\Omega^\omega S]$. □

Since the weights in the above experiment of Figure 2 are symmetric, this experiment also provides an example of the Riemannian gradient S-flow induced by the potential (3.56). Next, consider

$$L_{\mathcal{G}} = I_n - \Omega^\omega, \tag{3.57}$$

where $I_n \in \mathbb{R}^{n\times n}$ is the identity matrix. Since $I_n = \operatorname{Diag}(\Omega^\omega \mathbb{1}_n)$ by (2.28) is the degree matrix of the symmetric averaging matrix Ω^ω, $L_{\mathcal{G}}$ can be regarded as Laplacian (matrix) of the underlying undirected weighted graph $\mathcal{G} = (\mathcal{I}, \mathcal{E})$[1]. For the analysis of the S-flow, it will be convenient to rewrite the potential (3.56) accordingly.

[1]For undirected graphs, the graph Laplacian is commonly defined by the weighted *adjacency* matrices with diagonal entries 0, whereas $\Omega^\omega_{ii} = \omega_{ii} > 0$. The diagonal entries do not affect the quadratic form (3.58), however.

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

Proposition 3.10 *Under the assumption of Proposition 3.9, the potential (3.56) can be written in the form*

$$J(S) = \frac{1}{2}\langle S, L_{\mathcal{G}} S\rangle - \frac{1}{2}\|S\|^2 = \frac{1}{4}\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{N}_i}\omega_{ij}\|S_i - S_j\|^2 - \frac{1}{2}\|S\|^2. \tag{3.58}$$

The matrix $L_{\mathcal{G}}$ is symmetric, positive semidefinite and $L_{\mathcal{G}}\mathbb{1}_n = 0$.

Proof We have $J(S) = -\frac{1}{2}\langle S, (\Omega^{\omega} - I_n)S\rangle + \langle S, S\rangle = \frac{1}{2}(\langle S, L_{\mathcal{G}}S\rangle - \|S\|^2)$. Thus, we focus on the sum of (3.58).

First, note that $\|S_j - S_i\|^2 = \langle S_j, S_j - S_i\rangle + \langle S_i, S_i - S_j\rangle$. Since $\psi_{\mathcal{N}_i}(j) = \psi_{\mathcal{N}_j}(i)$ and $\omega_{ij} = \omega_{ji}$ by assumption, we have

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{N}_i}\omega_{ij}\langle S_j, S_j - S_i\rangle = \sum_{i,j\in\mathcal{I}}\psi_{\mathcal{N}_i}(j)\omega_{ij}\langle S_j, S_j - S_i\rangle = \sum_{i,j\in\mathcal{I}}\psi_{\mathcal{N}_j}(i)\omega_{ji}\langle S_j, S_j - S_i\rangle \tag{3.59a}$$

$$= \sum_{j\in\mathcal{I}}\sum_{i\in\mathcal{N}_j}\omega_{ji}\langle S_j, S_j - S_i\rangle = \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{N}_i}\omega_{ij}\langle S_i, S_i - S_j\rangle, \tag{3.59b}$$

where the last equality follows by renaming the indices i and j. Thus, using (2.28),

$$\begin{aligned}\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{N}_i}\omega_{ij}\|S_i - S_j\|^2 &= \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{N}_i}\omega_{ij}\langle S_j, S_j - S_i\rangle + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{N}_i}\omega_{ij}\langle S_i, S_i - S_j\rangle \\ &= 2\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{N}_i}\omega_{ij}\langle S_i, S_i - S_j\rangle = 2\sum_{i\in\mathcal{I}}\left\langle S_i, S_i - \sum_{j\in\mathcal{N}_i}\omega_{ij}S_j\right\rangle \\ &= 2\sum_{i\in\mathcal{I}}\langle S_i, (L_{\mathcal{G}}S)_i\rangle = 2\langle S, L_{\mathcal{G}}S\rangle.\end{aligned} \tag{3.60}$$

The properties of $L_{\mathcal{G}}$ follow from the symmetry of Ω^{ω}, nonnegativity of the quadratic form (3.60) and definition (3.57). □

4 Continuous-domain variational approach

In this section, we study a continuous-domain variational formulation of the potential of Proposition 3.10. We confine ourselves to uniform weights (2.28) and neighbourhoods (2.7) that only contain the nearest neighbours of each vertex i, such that $L_{\mathcal{G}}$ becomes the discretised ordinary Laplacian. As a result, we consider the problem to minimise the functional

$$E_{\alpha} : H^1(\mathcal{M}; \mathbb{R}^c) \to \mathbb{R}, \tag{4.1a}$$

$$E_{\alpha}(S) := \int_{\mathcal{M}} \|DS(x)\|^2 - \alpha\|S(x)\|^2 \mathrm{d}x, \qquad \alpha > 0. \tag{4.1b}$$

Throughout this section, $\mathcal{M} \subset \mathbb{R}^2$ is a simply connected bounded open subset in the Euclidean plane. Parameter α controls the interaction between regularisation and enforcing integrality when $S(x)$, $x \in \mathcal{M}$ is restricted to values in the probability simplex.

We prove well-posedness of problem (4.1) in Section 4.1 and consider Dirichlet boundary conditions in Section 4.2. In the former case, the set of minimisers is stated explicitly. The gradient descent flow corresponding to the latter case, initialised by means of given data and with parameter value $\alpha = 1$, may be seen as continuous-domain extension of the assignment flow,

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

that is parametrised according to Proposition 3.6 and operates at a small spatial scale in terms of the size $|\mathcal{N}_i|$ of uniform neighbourhoods (2.7) in the discrete formulation (2.32). We illustrate this by a numerical example (Section 4.3), based on discretising (4.1) and applying an algorithm that mimics the S-flow and converges to a local minimum of the non-convex functional (4.1), by solving a sequence of convex programs.

We point out that $\mathcal{M}$ could be turned into a Riemannian manifold using a metric that reflects images features (edges, etc.), as was proposed with the Laplace–Beltrami framework for image denoising [20]. In this work, however, we focus on the essential point that distinguishes image *denoising* from image *labeling*, that is, the interaction of the two terms (4.1) that essentially is a consequence of the information geometry of the assignment manifold $\mathcal{W}$ (2.21).

4.1 Well-posedness

Based on (2.3), we define the closed convex set

$$\mathscr{D}^1(\mathcal{M}) := \{S \in H^1(\mathcal{M}; \mathbb{R}^c) \colon \ S(x) \in \Delta_c \text{ a.e. in } \mathcal{M}\} \tag{4.2}$$

and focus on the variational problem

$$\inf_{S \in \mathscr{D}^1(\mathcal{M})} E_\alpha(S), \tag{4.3}$$

with E_α given by (4.1). E_α is smooth but non-convex. We specify the set of minimisers (Proposition 4.2). Recall notation (2.1).

Lemma 4.1 *Let $p \in \Delta^c$. Then, $\|p\| = 1$ if and only if $p \in \mathcal{B}_c$.*

Proof The 'if' statement is obvious. As for the 'only if', suppose $p \notin \mathcal{B}_c$, that is, $p_i < 1$ for all $i \in [c]$. Then, $p_i^2 < p_i$ and $\|p\|^2 < \|p\|_1 = 1$. □

Proposition 4.2 *The functional $E_\alpha \colon \mathscr{D}^1(\mathcal{M}) \to \mathbb{R}$ given by (4.1) is lower bounded,*

$$E_\alpha(S) \geq -\alpha \operatorname{Vol}(\mathcal{M}) > -\infty, \qquad \forall S \in \mathscr{D}^1(\mathcal{M}). \tag{4.4}$$

This lower bound is attained at some point in

$$\operatorname*{arg\,min}_{S \in \mathscr{D}^1(\mathcal{M})} E_\alpha(S) = \begin{cases} \{S_{e_1}, \ldots, S_{e_c}\}, & \text{if } \alpha > 0, \\ \{S_p \colon \mathcal{M} \to \Delta \colon p \in \Delta\}, & \text{if } \alpha = 0, \end{cases} \tag{4.5}$$

where, for any $p \in \Delta$, S_p denotes the constant vector field $x \mapsto S_p(x) = p$.

Proof Let $p \in \Delta$. Then, $\|p\|^2 \leq \|p\|_1 = 1$. It follows for $S \in \mathscr{D}^1(\mathcal{M})$ that

$$E_\alpha(S) \geq -\alpha \|S\|_{\mathcal{M}}^2 \geq -\alpha \|1\|_{\mathcal{M}} = -\alpha \operatorname{Vol}(\mathcal{M}), \tag{4.6}$$

which is (4.4).

We next show that the right-hand side of (4.5) specifies minimisers of E_α. For any $p \in \Delta$, the constant vector field S_p is contained in $\mathscr{D}^1(\mathcal{M})$. Consider specifically S_{e_i}, $i \in [c]$. Since $\|S_{e_i}(x)\| = \|e_i\| = 1$ and $DS_{e_i} \equiv 0$, the lower bound is attained, $E_\alpha(S_{e_i}) = -\alpha \operatorname{Vol}(\mathcal{M})$, and the

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

functions $\{S_{e_1}, \dots, S_{e_c}\}$ minimise E_α, for every $\alpha \geq 0$. If $\alpha = 0$, then the constant functions S_p are minimisers as well, for any $p \in \Delta$, since then

$$E_\alpha(S_p) = \|DS_p\|_{\mathcal{M}}^2 = 0 = -0 \cdot \operatorname{Vol}(\mathcal{M}). \tag{4.7}$$

We conclude by showing that no minimisers other than (4.5) exist. Let $S_* \in \mathscr{D}^1(\mathcal{M})$ be another minimiser of E_α with $E_\alpha(S_*) = -\alpha \operatorname{Vol}(\mathcal{M})$. We distinguish the two cases $\alpha = 0$ and $\alpha > 0$.

If $\alpha = 0$, then S_* satisfies (4.7) and $\|DS_*\|_{\mathcal{M}}^2 = 0$. Since $\|DS_{*;i}\|_{\mathcal{M}} \leq \|DS_*\|_{\mathcal{M}} = 0$ for every $i \in [c]$, S_* is constant by Lemma 2.3, that is, a $p \in \Delta$ exists such that $S_* = S_p$ a.e.

If $\alpha > 0$, then using the equation $E_\alpha(S_*) = -\alpha \operatorname{Vol}(\mathcal{M})$ and $\|S_*(x)\|^2 \leq 1$ gives

$$\alpha \operatorname{Vol}(\mathcal{M}) \leq \|DS_*\|_{\mathcal{M}}^2 + \alpha \operatorname{Vol}(\mathcal{M}) = \|DS_*\|_{1;\mathcal{M}}^2 - E_\alpha(S_*) = \alpha \|S_*\|_{\mathcal{M}}^2 \tag{4.8a}$$

$$\leq \alpha \|1\|_{\mathcal{M}} = \alpha \operatorname{Vol}(\mathcal{M}), \tag{4.8b}$$

which shows $\|DS_*\|_{\mathcal{M}} = 0$ and hence by Lemma 2.3 again $S_* = S_p$ for some $p \in \Delta$. The preceding inequalities also imply $\operatorname{Vol}(\mathcal{M}) = \|S_*\|_{\mathcal{M}}^2$, that is, $\|S_*(x)\| = 1$ a.e. By Lemma 4.1, we conclude $S_* = S_p$ with $p \in \mathcal{B}_c$, that is, $S_* \in \{S_{e_1}, \dots, S_{e_c}\}$. □

Proposition 4.2 highlights the effect of the concave term of the objective E_α (4.1): labellings are enforced in the absence of data. Below, the latter are taken into account (i) by imposing non-zero boundary conditions and (ii) by initialising a corresponding gradient flow (Section 4.3).

4.2 Fixed boundary conditions

In this section, we consider the case where boundary conditions are imposed by restricting the feasible set of problem (4.3) to

$$\mathcal{A}_g^1(\mathcal{M}) := \{S \in \mathcal{D}^1(\mathcal{M}) \colon S - g \in H_0^1(\mathcal{M}; \mathbb{R}^c)\} = \big(g + H_0^1(\mathcal{M}; \mathbb{R}^c)\big) \cap \mathcal{D}^1(\mathcal{M}), \tag{4.9}$$

for some fixed g that prescribes simplex-valued boundary values (in the trace sense). As intersection of a closed affine subspace and a closed convex set, $\mathcal{A}_g^1(\mathcal{M})$ is closed convex. A straightforward and reasonable choice of g is the given data $L_i(\mathbb{1}_S)$ lifted to the assignment manifold by means of the likelihood map (2.27), at boundary points $i \in \mathcal{I}$. We point out that besides making the problem well-posed, including the subproblems of the algorithm introduced in Section 4.3, the boundary condition has only a minor effect on local minimisers, as will be demonstrated numerically below. Employing Neumann boundary conditions is also a viable choice.

Weak lower semicontinuity is a key property for proving the existence of minimisers. In the case of E_α (4.1) this is not immediate, due to the lack of convexity.

Proposition 4.3 *The functional E_α given by (4.1) is weak sequentially lower semicontinuous on $\mathcal{A}_g^1(\mathcal{M})$, that is, for any sequence $(S_n)_{n\in\mathbb{N}} \subset \mathcal{A}_g^1(\mathcal{M})$ weakly converging to $S \in \mathcal{A}_g^1(\mathcal{M})$, the inequality*

$$E_\alpha(S) \leq \liminf_{n\to\infty} E_\alpha(S_n) \tag{4.10}$$

holds.

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

Proof Let $S_n \rightharpoonup S$ converge weakly in $\mathcal{A}_g^1(\mathcal{M}) \subset H_0^1(\mathcal{M};\mathbb{R}^c)$. Then, by Proposition 2.4(c),

$$\|S\|_{1;\mathcal{M}} \leq \liminf_{n\to\infty} \|S_n\|_{1;\mathcal{M}}. \tag{4.11}$$

Since $S, S_n \in \mathcal{A}_g^1(\mathcal{M})$, we also have $(S_n - g) \rightharpoonup (S-g)$ in $H_0^1(\mathcal{M};\mathbb{R}^c)$ by (4.9) and consequently $S_n \to S$ strongly in $L^2(\mathcal{M};\mathbb{R}^c)$ due to (2.38). Taking into account (4.11) and $\liminf_{n\to\infty} \|S_n\|_{\mathcal{M}} = \lim_{n\to\infty} \|S_n\|_{\mathcal{M}} = \|S\|_{\mathcal{M}}$, we obtain

$$E_\alpha(S) = \|S\|_{1;\mathcal{M}}^2 - (1+\alpha)\|S\|_{\mathcal{M}}^2 \leq \liminf_{n\to\infty} \|S_n\|_{1;\mathcal{M}}^2 + \liminf_{n\to\infty} \big(-(1+\alpha)\|S_n\|_{\mathcal{M}}^2\big) \tag{4.12a}$$

$$\leq \liminf_{n\to\infty} E_\alpha(S_n). \tag{4.12b}$$

□

We are now prepared to show that E_α attains its minimal value on $\mathcal{A}_g^1(\mathcal{M})$, following the basic proof pattern of [32, Chapter 38].

Theorem 4.4 *Let E_α be given by (4.1). There exists $S_* \in \mathcal{A}_g^1(\mathcal{M})$ such that*

$$E_\alpha^* := E_\alpha(S_*) = \inf_{S\in\mathcal{A}_g^1(\mathcal{M})} E_\alpha(S). \tag{4.13}$$

Proof Let $(S_n)_{n\in\mathbb{N}} \subset \mathcal{A}_g^1(\mathcal{M})$ be a minimising sequence such that

$$\lim_{n\to\infty} E_\alpha(S_n) = E_\alpha^*. \tag{4.14}$$

Then there exists some sufficiently large $n_0 \in \mathbb{N}$ such that

$$1 + E_\alpha^* \geq E_\alpha(S_n) = \|S_n\|_{1;\mathcal{M}}^2 - (1+\alpha)\|S_n\|_{\mathcal{M}}^2, \qquad \forall n \geq n_0. \tag{4.15}$$

Since $S_n(x) \in \Delta$ for a.e. $x \in \mathcal{M}$, we have $\|S_n\|_{\mathcal{M}}^2 \leq \mathrm{Vol}(\mathcal{M})$ and hence obtain

$$\|S_n\|_{1;\mathcal{M}}^2 \leq 1 + E_\alpha^* + (1+\alpha)\|S_n\|_{\mathcal{M}}^2 \leq 1 + E_\alpha^* + (1+\alpha)\,\mathrm{Vol}(\mathcal{M}), \qquad \forall n \geq n_0. \tag{4.16}$$

Thus, the sequence $(S_n)_{n\in\mathbb{N}} \subset H^1(\mathcal{M};\mathbb{R}^c)$ is bounded and, by Proposition 2.4(b), we may extract a weakly converging subsequence $S_{n_k} \rightharpoonup S_* \in H^1(\mathcal{M};\mathbb{R}^c)$. Since $\mathcal{A}_g^1(\mathcal{M}) \subset H^1(\mathcal{M};\mathbb{R}^c)$ is closed convex, Proposition 2.4(a) implies $S_* \in \mathcal{A}_g^1(\mathcal{M})$. Consequently, by Proposition 4.3 and (4.14),

$$E_\alpha(S_*) \leq \liminf_{k\to\infty} E_\alpha(S_{n_k}) = \lim_{k\to\infty} E_\alpha(S_{n_k}) = E_\alpha^*, \tag{4.17}$$

which implies $E_\alpha(S_*) = E_\alpha^*$, that is, $S_* \in \mathcal{A}_g^1(\mathcal{M})$ minimises E_α. □

4.3 Numerical algorithm and example

We consider the variational problem (4.13)

$$\inf_{S\in\mathcal{A}_g^1(\mathcal{M})} \int_{\mathcal{M}} \|DS\|^2 - \alpha\|S\|^2 dx, \tag{4.18}$$

for some fixed g specifying the boundary values $S|_{\partial\mathcal{M}} = g|_{\partial\mathcal{M}}$, and the problem to compute a local minimum numerically using an optimisation scheme that mimics the S-flow of Proposition 3.6.

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

Based on (4.9), we rewrite the problem in the form

$$\inf_{f\in H_0^1(\mathcal{M};\mathbb{R}^c)} \left\{ \|D(g+f)\|_{\mathcal{M}}^2 - \alpha\|g+f\|_{\mathcal{M}}^2 + \delta_{\mathcal{D}^1(\mathcal{M})}(g+f) \right\} \tag{4.19a}$$

$$= \inf_{f\in H_0^1(\mathcal{M};\mathbb{R}^c)} \left\{ \|Df\|_{\mathcal{M}}^2 + 2\langle Dg, Df\rangle_{\mathcal{M}} - \alpha\big(\|f\|_{\mathcal{M}}^2 + 2\langle g,f\rangle_{\mathcal{M}}\big) + \delta_{\mathcal{D}^1(\mathcal{M})}(g+f) \right\} + C, \tag{4.19b}$$

where the last constant C collects terms not depending on f. We discretise the problem as follows. f becomes a vector $f\in\mathbb{R}^{c\,n}$ with $n=|\mathcal{I}|$ subvectors $f_i\in\mathbb{R}^c$, $i\in\mathcal{I}$ or alternatively with $c=|\mathcal{J}|$ subvectors f^j, $j\in\mathcal{J}$. The inner product $\langle g, f\rangle_{\mathcal{M}}$ is replaced by $\langle g, f\rangle = \sum_{i\in[n]}\langle g_i, f_i\rangle = \sum_{j\in[c]}\langle g^j, f^j\rangle$. We keep the symbols f, g for simplicity and indicate the discretised setting by the subscript n as introduced next.

D becomes a gradient matrix D_n that estimates the gradient of each subvector f^j separately, such that

$$L_n f := D_n^\top D_n f \tag{4.20}$$

is the basic discrete five-point stencil Laplacian applied to each subvector f^j. The feasible set $\mathcal{D}^1(\mathcal{M})$ is replaced by the closed convex set

$$\mathcal{D}_n := \{f\geq 0\colon\ \langle \mathbb{1}_c, f_i\rangle = 1,\ \forall i\in\mathcal{I}\}. \tag{4.21}$$

Thus the discretised problem reads

$$\inf_f \left\{ \|D_n f\|^2 + 2\langle L_n g - \alpha g, f\rangle - \alpha\|f\|^2 + \delta_{\mathcal{D}_n}(g+f) \right\}. \tag{4.22}$$

Having computed a local minimum f_*, the corresponding local minimum of (4.18) is $S_* = g + f_*$.

In order to compute f_*, we applied the proximal forward–backward scheme

$$f^{(k+1)} = \arg\min_f \left\{ \|D_n f\|^2 + 2\langle L_n g - \alpha(g+f^{(k)}), f\rangle + \frac{1}{2\tau_k}\|f-f^{(k)}\|^2 + \delta_{\mathcal{D}_n}(g+f) \right\}, \quad k\geq 0, \tag{4.23}$$

with proximal parameters τ_k, $k\in\mathbb{N}$ and initialisation $f_i^{(0)}$, $i\in\mathcal{I}$ specified further below. The iterative scheme (4.23) is a special case of the PALM algorithm [8, Section 3.7]. Ignoring the proximal term, each problem (4.23) amounts to solve c (discretised) Dirichlet problems with the boundary values of g^j, $j\in[c]$ imposed, and with right-hand sides that change during the iteration since they depend on $f^{(k)}$. The solutions $(f^j)^{(k)}$, $j\in\mathcal{J}$ to these Dirichlet problems depend on each other, however, through the feasible set (4.21). At each iteration k, problem (4.23) can be solved by convex programming. The proximal parameters τ_k act as step sizes such that the sequence $f^{(k)}$ does not approach a local minimum too rapidly. Then the interplay between the linear form that adapts during the iteration and the regularising effect of the Laplacians can find a labelling (partition) corresponding to a good local optimum.

As for g, we chose $g_i = L_i(\mathbb{1}_{\mathcal{S}})$, $i\in\mathcal{I}$ at boundary vertices i and $g_i = 0$ at every interior vertex i. Consequently, with the initialisation $f_i^{(0)} = L_i(\mathbb{1}_{\mathcal{S}})$, $i\in\mathcal{I}$ at interior vertices (the boundary values of f are zero), the sequence $S^{(k)} = g + f^{(k)}$ mimics the S-flow of Proposition 3.6 where the given data also show up in the initialisation $\overline{S}(0)$ only.

Figure 3 provides an illustration using the experimental setup from Figure 2. Comparing the result depicted by Figure 3 with Figure 2 confirms that the continuous-domain formulations

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

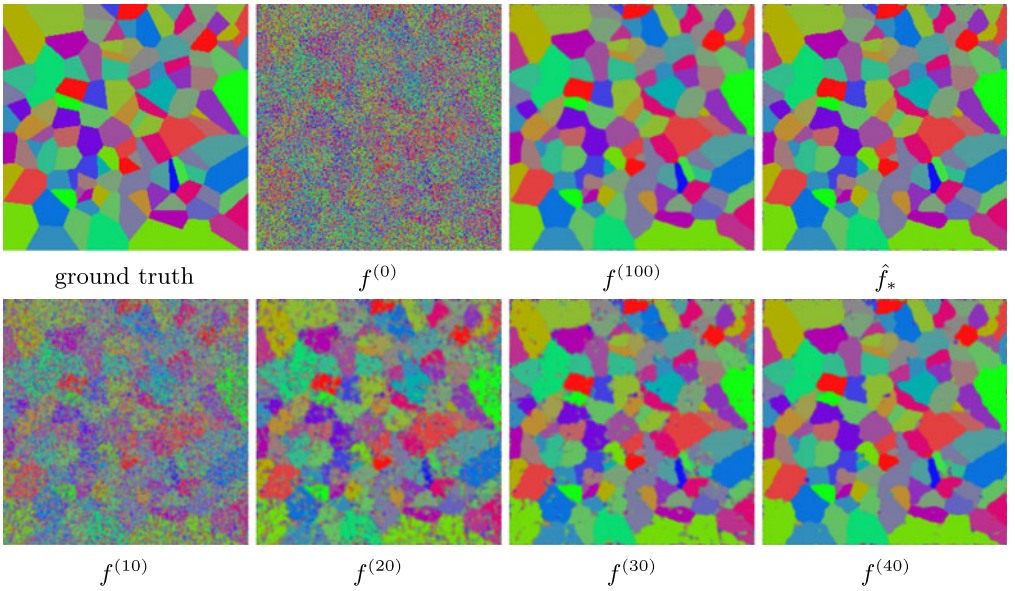

FIGURE 3. Evaluation of the numerical scheme (4.23) that mimics the S-flow of Proposition 3.6. Parameter values: $\alpha = 1$ and $\tau_k = \tau = 10$, $\forall k$. Top, from left to right: Ground truth, noisy input data $f^{(0)}$, iterate $f^{(100)}$ and the approximate local minimiser $\hat{f}_*$ resulting from $f^{(100)}$ by a trivial rounding step. $S^{(k)} = f^{(k)} + g$ differs from $f^{(k)}$ by the boundary values corresponding to the noisy input data. Inspecting the values of $f^{(100)}$ close to the boundary shows that the influence of boundary noise is minimal. Bottom, from left to right: The iterates $f^{(10)}$, $f^{(20)}$, $f^{(30)}$, $f^{(40)}$. Taking into account rounding as post-processing step, the sequence $f^{(k)}$ quickly converges after rounding to a reasonable partition. About 50 more iterations are required to fix the values at merely few hundred remaining pixels. Slight rounding of the geometry of the components of the partition, in comparison to ground truth, corresponds to using uniform weights (2.28) for the assignment flow.

discussed above represents the assignment flow at a small spatial scale. The two numerical approaches are quite different, however. While geometric numerical integration employs many steps with only local explicit interactions of variables, the proximal iteration (4.23) entails global communication of the variables at each iteration through solving a linear PDE together with a pointswise simplex constraint. As a result, a one order of magnitude smaller number of (more expensive) iterations is required to converge.

4.4 A PDE characterising optimal assignment flows

Proposition 4.5 *Let S_* solve the variational problem (4.18). Then, S_* satisfies the variational inequality*

$$\langle DS_*, DS - DS_* \rangle_{\mathcal{M}} - \alpha \langle S_*, S - S_* \rangle_{\mathcal{M}} \geq 0, \qquad \forall S \in \mathcal{A}_g^1(\mathcal{M}). \tag{4.24}$$

Proof Functional E_α given by (4.18) is Gateaux-differentiable with derivative

$$\langle E_\alpha'(S_*), S \rangle_{H^{-1}(\mathcal{M};\mathbb{R}^c) \times H_0^1(\mathcal{M};\mathbb{R}^c)} = 2\big(\langle DS_*, S \rangle_{\mathcal{M}} - \alpha \langle S_*, S \rangle_{\mathcal{M}} \big). \tag{4.25}$$

The assertion follows from applying Theorem 2.5. □

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

We conclude this section by deriving a PDE corresponding to (4.24), that a minimiser S_* is supposed to satisfy in the weak sense. The derivation is *formal* in that we adopt the unrealistic regularity assumption

$$S_* \in \mathcal{A}_g^2(\mathcal{M}), \tag{4.26}$$

with $\mathcal{A}_g^2(\mathcal{M})$ defined analogous to (4.9). While this will hold for the continuous-domain *linear* problems corresponding to (4.23) at each step k of the iteration and for sufficiently smooth $\partial\mathcal{M}$, it will not hold in the limit $k \to \infty$, since we expect (and wish) S_* to become discontinuous, contrary to the regularity assumption (4.26) and the continuity implied by the Sobolev embedding theorem for $\mathcal{M} \subset \mathbb{R}^d$ with $d = 2$. Nevertheless, since the PDE provides another interpretation of the assignment flow, we state it – see (4.32) below – and hope it will stimulate further research.

In view of the assumption (4.26), set

$$S_* = g + f_*, \qquad f_* \in H_0^2(\mathcal{M}; \mathbb{R}^c). \tag{4.27}$$

Inserting S_* and $S = g + h,\ h \in H_0^1(\mathcal{M}; \mathbb{R}^c)$, into (4.24) and partial integration, gives

$$\langle -\Delta S_* - \alpha S_*, h - f_* \rangle_{\mathcal{M}} \geq 0, \tag{4.28}$$

where $\Delta S_* = (\Delta S_{*;1}, \ldots, \Delta S_{*;c})^\top$ applies componentwise. Using the shorthands

$$\nu_\alpha(S_*) = -\Delta S_* - \alpha S_*, \tag{4.29a}$$

$$\mu_\alpha(S_*) = \nu_\alpha(S_*) - \langle \nu_\alpha(S_*), S_* \rangle_{\mathbb{R}^2} \mathbb{1}_c, \tag{4.29b}$$

where $\langle \nu_\alpha(S_*), S_* \rangle_{\mathbb{R}^2}$ denotes the function $x \mapsto \big\langle \nu_\alpha(S_*)(x), S_*(x) \big\rangle$, $x \in \mathcal{M}$, we have

$$\langle \mu_\alpha(S_*), S_* \rangle_{\mathcal{M}} = 0, \tag{4.30a}$$

since $\langle \mathbb{1}_c, S_*(x) \rangle = 1$ for a.e. x, and

$$\langle \mu_\alpha(S_*), S \rangle_{\mathcal{M}} = \langle \nu_\alpha(S_*), h - f_* \rangle_{\mathcal{M}} \geq 0, \tag{4.30b}$$

which is (4.28). Since $S(x) \geq 0$ a.e. in $\mathcal{M}$ and may have arbitrary support, we deduce from the inequality $\langle \mu_\alpha(S_*), S \rangle_{\mathcal{M}} \geq 0$ and from the self-duality of the nonnegative orthant $\mathbb{R}_+^c$ that $\mu_\alpha(S_*) \geq 0$ a.e. in $\mathcal{M}$. Since also $S_* \geq 0$ a.e., this implies that equation (4.30a) holds pointwise a.e. in $\mathcal{M}$:

$$\mu_\alpha(S_*)(x) S_*(x) = \nu_\alpha(S_*)(x) S_*(x) - \big\langle \nu_\alpha(S_*)(x) S_*(x) \big\rangle S_*(x) = 0 \qquad \text{a.e. in } \mathcal{M}. \tag{4.31}$$

Substituting $\nu_\alpha(S_*)$, we deduce that a minimiser $S_* = g + f_*$ characterised by the variational inequality (4.24) weakly satisfies the PDE

$$R_{S_*}(-\Delta S_* - \alpha S_*) = 0, \tag{4.32}$$

where R_{S_*} defined by (2.16) applies $R_{S_*(x)}$ to vector $(-\Delta S_* - \alpha S_*)(x)$ at every $x \in \mathcal{M}$.

Remark 4.6 (Comments)

(1) *We point out that computing a vector field S_* satisfying (4.24) is difficult in practice, due to the nonconvexity of problem (4.18). On the other hand, the algorithm proposed in Section 4.3*

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

in the result illustrated by Figure 3 shows that good suboptima can be computed by merely solving a sequence of simple problems.

(2) *As already pointed out at the beginning of this section, the derivation of the PDE (4.32) is merely a formal one, due to the unrealistic regularity assumption (4.26). While the numerical result (Figure 3) clearly reflects the desired piecewise constancy and discontinuity of S_*, this property conflicts with assumption (4.26).*

5 Conclusion

We presented a novel parametrisation of the assignment flow for contextual data classification on graphs. The dominating part of the flow admits the interpretation as Riemannian gradient flow with respect to the underlying information geometry, unlike the original formulation of the assignment flow. A decomposition of the corresponding potential by means of a non-local graph Laplacian makes explicit the interaction of two processes: regularisation of label assignments and gradual enforcement of unambiguous decisions. The assignment flow combines these aspects in a seamless way, unlike traditional approaches where solutions to convex relaxations require postprocessing. It is remarkable that this behaviour is solely induced by the underlying information geometry.

We studied a continuous-domain variational formulation as counterpart of the discrete formulation restricted to a local discrete Laplacian (nearest neighbour interaction). A numerical algorithm in terms of a sequence of simple linear elliptic problems reproduces results that were obtained with the original formulation of the assignment flow using completely different numerics (geometric ODE integration). This illustrates the derived mathematical relations.

We outline three attractive directions of further research.

- We clarified in Section 4 that the inherent *smooth* setting of the assignment flow (2.32) translates under suitable assumptions to the sequence of linear (discretised) elliptic PDE problems (4.23) together with a simple convex constraint. We did not touch on the limit problem, however. More mathematical work is required here, cf. Remark 4.6.

  Since the assignment flow returns image *partitions* when applied to image features on a grid graph, the situation reminds us of the Mumford–Shah functional [23] – more precisely: its restriction to piecewise constant functions – and its approximation by a sequence of Γ-converging smooth elliptic problems [4]. Likewise, one may regard the concave second term of (4.18) together with the convex constraint $S \in \mathcal{A}_g^1$ as a vector-valued counterpart of the basic nonnegative double-well potential of scalar phase-field models for binary segmentation [28, 12]. In these works, too, non-smooth limit cases result from Γ-converging simpler problems.
- Adopting the viewpoint of evolutionary dynamics [17] on label assignment, the assignment flow may be characterised as spatially coupled replicator dynamics. To the best of our knowledge, our paper [6] seems to be the first one that used information theory to formulate this spatial coupling. Some consequences of the geometry were elaborated in the present paper and discussed above.

  We point out that the literature on evolutionary dynamics in general, and specifically on the replicator equation, is vast. We merely point out few works on models involving the replicator equation and spatial interaction in physics [29, 13], applied mathematics [24, 9], including

https://doi.org/10.1017/S0956792520000273 Published online by Cambridge University Press

extensions to scenarios with an infinite number of strategies (as opposed to selecting from a finite set of labels) – see [3] and references therein.

In this context, our work might stimulate researchers working on *spatially extended* evolutionary dynamics in various scientific disciplines. In particular, generalising our approach to continuous-domain integro-differential models seems attractive that conform to the assignment flow with *non-local* interactions (i.e. with larger neighbourhoods $|\mathcal{N}_i|$, $i \in \mathcal{I}$) and the underlying geometry.

- Last but not least, our work may support a better understanding of *learning* with networks. Our preliminary work on learning the weights (2.28) using the *linearised* assignment flow [18] on a *single* graph ('layer') revealed the model expressiveness of this limited scenario, on the one hand, and that subdividing complex learning tasks in this way avoids 'black box behaviour', on the other hand. We hope that the continuous-domain perspective developed in this paper in terms of sequences of linear PDEs will support our further understanding of learning with hierarchical 'deeper' architectures.

Acknowledgements

Financial support by the German Science Foundation (DFG), grant GRK 1653, is gratefully acknowledged. This work was supported by Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy EXC-2181/1 – 390900948 (the Heidelberg STRUCTURES Excellence Cluster).

Conflict of interest

The authors declare that there is no conflict of interest.

References

[1] Attouch, H., Buttazzo, G. & Michaille, G. (2014) *Variational Analysis in Sovolev and BV Spaces: Applications to PDEs and Optimization*, 2nd ed., SIAM, Philadelphia, PA.

[2] Amari, S.-I. & Nagaoka, H. (2000) *Methods of Information Geometry*, American Mathematical Society and Oxford University Press, Providence, Rhode Island.

[3] Ambrosio, L., Fornasier, M., Morandotti, M. & Savaré, G. (2018) *Spatially Inhomogeneous Evolutionary Games*, CoRR abs/1805.04027.

[4] Ambrosio, L. & Tortorelli, V. M. (1990) Approximation of functional depending on jumps by elliptic functional via Γ-convergence. *Comm. Pure Appl. Math.* **43**(8), 999–1036.

[5] Antun, V., Renna, F., Poon, C., Adcock, B. & Hansen, A. C. (2019) *On Instabilities of Deep Learning in Image Reconstruction - Does AI Come at a Cost?*, CoRR abs/1902.05300.

[6] Åström, F., Petra, S., Schmitzer, B. & Schnörr, C. (2017) Image labeling by assignment. *J. Math. Imaging Vis.* **58**(2), 211–238.

[7] Beck, A. & Teboulle, M. (2003) Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operat. Res. Lett.* **31**(3), 167–175.

[8] Bolte, J., Sabach, S. & Teboulle, M. (2014) Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Progr. Ser. A* **146**(1–2), 459–494.

[9] Bratus, A. S., Posvyanskii, V. P. & Novozhilov, A. S. (2014) Replicator equations and space. *Math. Model. Nat. Phenomena* **9**(3), 47–67.

[10] Chambolle, A., Cremers, D. & Pock, T. (2012) A convex approach to minimal partitions. *SIAM J. Imag. Sci.* **5**(4), 1113–1158.

[11] Coveney, P. V., Dougherty, E. R. & Highfield, R. R. (2016) Big data need big theory too. *Phil. Trans. R. Soc. Lond. A* **374**, 20160153.

[12] CRISTOFERI, R. & THORPE, M. (2020) Large data limit for a phase transition model with the p-Laplacian on point clouds. *Europ. J. Appl. Math.* **31**(2), 185–231.
[13] DEFOREST, R. & BELMONTE, A. (2013) Spatial pattern dynamics due to the fitness gradient flux in evolutionary games. *Phys. Rev. E* **87**(6), 062138.
[14] ELAD, M. (2017) Deep, deep trouble: deep learning's impact on image processing, mathematics, and humanity, *SIAM News* **50**(4).
[15] HABER, E. & RUTHOTTO, L. (2017) Stable architectures for deep neural networks. *Inverse Prob.* **34**(1), 014004.
[16] HE, K., ZHANG, X., REN, S. & SUN, J. (2016) Deep residual learning for image recognition. In: *Proceedings of CVPR*.
[17] HOFBAUER, J. & SIEGMUND, K. (2003) Evolutionary game dynamics. *Bull. Amer. Math. Soc.* **40**(4), 479–519.
[18] HÜHNERBEIN, R., SAVARINO, F., PETRA, S. & SCHNÖRR, C. (2019) Learning adaptive regularization for image labeling using geometric assignment. In: *Proceedings of SSVM*, Springer.
[19] JOST, J. (2017) *Riemannian Geometry and Geometric Analysis*, 7th ed., Springer-Verlag, Berlin, Heidelberg.
[20] KIMMEL, R., MALLADI, R. & SOCHEN, N. (2000) Images as embedded maps and minimal surfaces: movies, color, texture, and volumetric images. *Int. J. Comp. Vis.* **39**(2), 111–129.
[21] LELLMANN, J. & SCHNÖRR, C. (2011) Continuous multiclass labeling approaches and algorithms. *SIAM J. Imag. Sci.* **4**(4), 1049–1096.
[22] LIU, G.-H. & THEODOROU, E. A. (2019) *Deep Learning Theory Review: An Optimal Control and Dynamical Systems Perspective*, CoRR abs/1908.10920.
[23] MUMFORD, D. & SHAH, J. (1989) Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.* **42**, 577–685.
[24] NOVOZHILOV, A., POSVNOVOZH, V. P. & BRATUS, A. S. (2011) On the reaction–diffusion replicator systems: spatial patterns and asymptotic behaviour. *Russ. J. Numer. Anal. Math. Modell.* **26**(6), 555–564.
[25] ROCKAFELLAR, R. T. & WETS, R. J.-B. (2009) *Variational Analysis*, Vol. 317, Springer Science & Business Media, Berlin and Heidelberg, Germany.
[26] SANDHOLM, W. H. (2010) *Population Games and Evolutionary Dynamics*, MIT Press, Cambridge, Massachusetts.
[27] SCHNÖRR, C. (in press) Assignment flows. In: P. Grohs, M. Holler & A. Weinmann (editors), *Variational Methods for Nonlinear Geometric Data and Applications*, Springer, Cham, Switzerland.
[28] STERNBERG, P. (1991) Vector-valued local minimizers of nonconvex variational problems. *Rocky-Mountain J. Math.* **21**(2), 799–807.
[29] TRAULSEN, A. & CLAUSSEN, J. C. (2004) Similarity-based cooperation and spatial segregation. *Phys. Rev. E* **70**(4), 046128.
[30] WEINAN E. (2017) A proposal on machine learning via dynamical systems. *Comm. Math. Stat.* **5**(1), 1–11.
[31] WEINAN E., HAN, J. & LI, Q. (2019) A mean-field optimal control formulation of deep learning. *Res. Math. Sci.* **6**(10), 41 p.
[32] ZEIDLER, E. (1985) *Nonlinear Functional Analysis and its Applications*, Vol. 3, Springer, New York, United States.
[33] ZERN, A., ZEILMANN, A. & SCHNÖRR, C. (2020) *Assignment Flows for Data Labeling on Graphs: Convergence and Stability*, CoRR abs/2002.11571.
[34] ZERN, A., ZISLER, M., PETRA, S. & SCHNÖRR, C. (in press) Unsupervised assignment flow: Label learning on feature manifolds by spatially regularized geometric assignment. *J. Math. Imaging Vis*. https://doi.org/10.1007/s10851-019-00935-7.
[35] ZEILMANN, A., SAVARINO, F., PETRA, S. & SCHNÖRR, C. (2020) Geometric numerical integration of the assignment flow. *Inverse Prob.* **36**(3), 034004 (33pp).
[36] ZIEMER, W. P. (1989) *Weakly Differentiable Functions*, Springer, New York, United States.
[37] ZISLER, M., ZERN, A., PETRA, S. & SCHNÖRR, C. (2019) Unsupervised labeling by geometric and spatially regularized self-assignment. In: *Proceedings of SSVM*, Springer.