A Variational Perspective on the Assignment Flow

Fabrizio Savarino and Christoph Schnörr

Image and Pattern Analysis Group, Heidelberg University, Germany
fabrizio.savarino@iwr.uni-heidelberg.de
schnoerr@math.uni-heidelberg.de

Abstract. The image labeling problem can be described as assigning to each pixel a single element from a finite set of predefined labels. Recently, a smooth geometric approach for inferring such label assignments was proposed by following the Riemannian gradient flow of a given objective function on the so-called assignment manifold. Due to the specific Riemannian structure, this results in a coupled replicator dynamic incorporating local spatial geometric averages of lifted data-dependent distances. However, in this framework an approximation of the flow is necessary in order to arrive at explicit formulas. We propose an alternative variational model, where lifting and averaging are decoupled in the objective function so as to stay closer to established approaches and at the same time preserve the main ingredients of the original approach: the overall smooth geometric setting and regularization through geometric local averages. As a consequence the resulting flow is explicitly given, without the need for any approximation. Furthermore, there exists an interesting connection to graphical models.

Keywords: image labeling, assignment manifold, assignment flow, geometric optimization, Riemannian gradient flow, replicator equation, multiplicative updates

1 Introduction

Overview, Motivation. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph representing a certain spatial structure and denote by $f\colon \mathcal{V} \to \mathcal{F}$ some given data on that graph with values in a feature space $\mathcal{F}$. A *labeling* of f on $\mathcal{V}$ with predefined labels $\mathcal{L} = \{l_1, \ldots, l_n\}$ is a map $A\colon \mathcal{V} \to \mathcal{L}$ assigning to every vertex $i \in \mathcal{V}$ a label $A_i \in \mathcal{L}$. By identifying the nodes $\mathcal{V}$ with the numbers $\{1, \ldots, m\}$, for $m := |\mathcal{V}|$, a labeling A corresponds to a vector $A \in \mathcal{L}^m$. In the case of *image labeling*, the graph $\mathcal{G}$ might be a grid graph embedded into the image domain $\Omega \subset \mathbb{R}^2$ and f represents

Acknowledgments: We gratefully acknowledge support by the German Science Foundation, grant GRK 1653.

some observed raw image data $\mathcal{F} = [0,1]^3$ or some features extracted from the image by standard methods. Depending on the domain of application, one is usually interested in finding an optimal labeling with respect to a quality measure, called *objective function*. In general the task of computing globally optimal labels results in an NP-hard problem and therefore several relaxations are used to arrive at a computationally feasible formulation [7].

In [2] a new smooth geometric approach was suggested, which is modeled on the manifold of row-stochastic matrices with full support, called the *assignment manifold* and denoted by $\mathcal{W} \subset \mathbb{R}^{m \times n}$ (for details see Section 2). By choosing the Fisher-Rao (information) metric, $\mathcal{W}$ is turned into a Riemannian manifold. Their basic idea is to encode labelings as points on the assignment manifold, exploit the Riemannian setting for constructing an objective function $E\colon \mathcal{W} \to \mathbb{R}$ using Riemannian means and optimizing it by following the Riemannian gradient flow. After a simplifying assumption and approximating the Riemannian mean to first order by the geometric mean, they arrive at the following dynamical system, called *assignment flow*

$$\dot{W}(t) = \Pi_{W(t)} S(W(t)) \tag{1.1}$$

where S consists of certain geometric means and Π_W is a linear map (see Section 2).

While the overall geometric model constitutes an interesting new approach to the labeling problem and performs very well, there are some mathematical points to address: The relation to classical approaches with objective function $E = E_{\text{data}} + E_{\text{reg}}$ has not been worked out, where E_{data} is a data dependent and E_{reg} a regularization term. It can be shown, that there exists no potential of the vector field (1.1) in the Fisher-Rao geometry, which implies that the flow is not variational. Furthermore, the above mentioned simplifying assumption and approximation are unavoidable, since otherwise there is no closed form solution for the Riemannian mean.

Contribution. We propose a variational model where the lifting and averaging is decoupled in a way similar to more classical approaches of the form $E = E_{\text{data}} + E_{\text{reg}}$ mentioned above. In this alternative model we are able to completely avoid the need for any approximation and simplifying assumptions, while still exploiting the Riemannian structure of the setting. Additionally, there is an interesting connection to graphical models, a well established formulation of image labeling.

2 Preliminaries

Basic Notation. We assume $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is an *undirected* graph. If two nodes i and j are connected by an undirected edge $ij \in \mathcal{E}$ then we call i and j adjacent and denote this relation by $i \sim j$. The neighborhood of node i is the set $\mathcal{N}(i) := \{j \in \mathcal{V}\colon i \sim j\}$. The number of nodes will be denoted by $m := |\mathcal{V}|$ and the number of labels by $n := |\mathcal{L}|$. We use the abbreviation $[k] = \{1, 2, \ldots, k\}$ for $k \in \mathbb{N}$ and identify $[m]$ with $\mathcal{V}$ as well as $[n]$ with $\mathcal{L}$. For a matrix $M \in \mathbb{R}^{m \times n}$ we denote the i-th row of M as M_i. For any two vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n_{>0}$,

we denote the componentwise product and division by $xy = (x_1y_1, \dots, x_ny_n)^\top$ and $\frac{x}{y} = (\frac{x_1}{y_1}, \dots, \frac{x_n}{y_n})^\top$ respectively. If $g\colon \mathbb{R} \to \mathbb{R}$ is a scalar function, then $g(x)$ denotes the componentwise application of g, i.e. $e^x = (e^{x_1}, \dots, e^{x_n})^\top$. The standard basis of $\mathbb{R}^n$ is denoted by $\{e_1, \dots, e_n\}$ and the standard inner product on $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ respectively by $\langle \cdot, \cdot \rangle$. We set $\mathbb{1}_n = (1, 1, \dots, 1)^\top \in \mathbb{R}^n$.

The Assignment Manifold. We briefly introduce the necessary geometric setting of the assignment manifold from [2]. Let $\Delta_n = \{p \in \mathbb{R}^n : p_i \geq 0 \text{ for } i = 1, \dots, n\,, \langle p, \mathbb{1}\rangle = 1\}$ denote the probability simplex and $c := \frac{1}{n}\mathbb{1}_n$ the barycenter. The relative interior of Δ_n is given by

$$\mathcal{S} := \operatorname{rint}(\Delta_n) = \{p \in \Delta_n : p_i > 0 \text{ for } i = 1, \dots, n\} \tag{2.1}$$

and is a smooth manifold of dimension $n-1$ with a global chart and an $n-1$ dimensional constant tangent space

$$T_p\mathcal{S} = \{v \in \mathbb{R}^n : \langle v, \mathbb{1}\rangle = 0\} =: T \subset \mathbb{R}^n \quad \text{for all} \quad p \in \mathcal{S}. \tag{2.2}$$

The orthogonal projection of $\mathbb{R}^n$ to T with respect to the standard inner product is given by

$$P_T[x] := \big(I - \tfrac{1}{n}\mathbb{1}\mathbb{1}^\top\big)x \tag{2.3}$$

The *lifting map* $\exp\colon T\mathcal{S} = \mathcal{S} \times T \to \mathcal{S}$ is defined as

$$(p, u) \mapsto \exp_p(u) := \frac{pe^u}{\langle p, e^u\rangle}. \tag{2.4}$$

The map $\exp_p\colon T \to \mathcal{S}$ is a diffeomorphism for every $p \in \mathcal{S}$ with inverse $\exp_p^{-1}(q) = P_T \log(\frac{q}{p})$ and since $T \subset \mathbb{R}^n$ is a linear space, it can be used as a chart for $\mathcal{S}$. The lifting map can also be viewed as $\exp_p\colon \mathbb{R}^n \to \mathcal{S}$ with $\exp_p \circ P_T = \exp_p$, however, this is not an invertible map anymore.

The Fisher-Rao metric endows $\mathcal{S}$ with a Riemannian structure given by

$$g_p\colon T \times T \to \mathbb{R}, \quad g_p(u, v) = \langle u, \operatorname{Diag}(\tfrac{1}{p})v\rangle, \tag{2.5}$$

for $p \in \mathcal{S}$ and $u, v \in T$. Denote by $2\mathbb{S}^{n-1} \subset \mathbb{R}^n$ the sphere of radius 2 with Riemannian metric induced by the Euclidean inner product of $\mathbb{R}^n$. There is an isomorphism, called *sphere map* $\psi\colon \mathcal{S} \to 2\mathbb{S}^{n-1} \cap \mathbb{R}^n_{>0}$, given by $p \mapsto 2\sqrt{p}$ (cf. [2, Sec. 2.1]). Due to the form of ψ, the geometry can be continuously extended to Δ_n. As a consequence, the Riemannian distance between $p, q \in \mathcal{S}$ is given by

$$d_{\mathcal{S}}(p, q) = 2\arccos(\langle\sqrt{p}, \sqrt{q}\rangle) \leq \pi. \tag{2.6}$$

There is also an explicit formula for the exponential map of the Riemannian manifold Exp_p and its inverse (cf. [2, Prop 2 and eq. (7.16b)]), where the latter is given by

$$\operatorname{Exp}_p^{-1}(q) = \frac{d_{\mathcal{S}}(p, q)}{\sqrt{1 - \langle\sqrt{p}, \sqrt{q}\rangle^2}}\big(\sqrt{pq} - \langle\sqrt{p}, \sqrt{q}\rangle p\big). \tag{2.7}$$

For a scalar valued function $f\colon \mathcal{S} \to \mathbb{R}$, the Riemannian gradient (cf. [1]) at $p \in \mathcal{S}$ is the vector $\nabla_{\mathcal{S}} f(p) \in T$ uniquely characterized via the differential of f by $Df(p)[v] = g_p(\nabla_{\mathcal{S}} f(p), v)$ for all $v \in T$. Define the *replicator operator* as

$$\Pi_p\colon \mathbb{R}^n \to T \quad x \mapsto \Pi_p[x] := \big(\operatorname{Diag}(p) - pp^\top\big)x \quad \text{for} \quad p \in \mathcal{S}. \tag{2.8}$$

An elementary calculation shows that $\Pi_p \circ P_T = \Pi_p$. Viewed as a linear map $\Pi_p\colon T \to T$ an inverse exists and is given by $(\Pi_p)^{-1} = P_T \operatorname{Diag}(\frac{1}{p})$.

Denote by $\nabla f(p) \in T$ the Riemannian gradient of $\mathcal{S}$ with the standard inner product as Riemannian metric, then $\nabla_{\mathcal{S}} f$ and ∇f are connected by (cf. [2, Prop. 1])

$$\nabla_{\mathcal{S}} f(p) = p\nabla f(p) - \langle p, \nabla f(p)\rangle p = \Pi_p[\nabla f(p)]. \tag{2.9}$$

The Riemannian gradient flow on $\mathcal{S}$ can be transformed onto T by $\exp_c$. According to [9, Cor. 1 and Lem. 4] we have $t \mapsto p(t) \in \mathcal{S}$ with $p(0) = c$ solves gradient flow (2.10)(a) if and only if $t \mapsto v(t) \in T$ with $v(0) = 0$ and $p(t) = \exp_c(v(t))$ solves the gradient flow (2.10)(b)

$$\text{(a)} \quad \dot{p}(t) = \nabla_{\mathcal{S}} f(p(t)) \qquad \text{(b)} \quad \dot{v}(t) = \nabla f\big(\exp_c(v(t))\big) \tag{2.10}$$

The *assignment manifold* $\mathcal{W}$ is defined to be the product manifold $\mathcal{W} := \mathcal{S}^m$ with tangent space given by $T_W\mathcal{W} = T^m =: \mathcal{T}$ for $W \in \mathcal{W}$. The Fisher Rao metric on $\mathcal{S}$ induces a Riemannian metric on $\mathcal{W}$ via the product metric, thus the above formulas carry over to the product manifold setting by applying them componentwise. In the following we use the description of $\mathcal{W}$ as a set of matrices $\mathcal{W} = \{W \in \mathbb{R}^{m\times n}_{>0}\colon W\mathbb{1} = \mathbb{1}\}$ together with $\mathcal{T} = \{V \in \mathbb{R}^{m\times n}\colon V\mathbb{1} = 0\}$.

3 Model

Motivation: The Assignment Flow. In order to better motivate the form of our variational approach below and the parallels to [2], we review the core concept of the assignment flow in a bit more detail. The basic idea of modeling the labeling problem on the assignment manifold is to encode label $l_j \in \mathcal{L}$ by the j-th standard basis vector e_j of $\mathbb{R}^n$, which is a corner of Δ_n. With this, a labeling $A \in \mathcal{L}^m$ corresponds to an assignment matrix $W \in \Delta_n$ with i-th row $W_i = e_j$ if $A_i = l_j$. These integral labelings are relaxed to the assignment manifold $\mathcal{W}$ by allowing them to be fully probabilistic. Let $f\colon \mathcal{V} \to \mathcal{F}$ be some given data on the graph $\mathcal{G}$ with values in a feature space $\mathcal{F}$. Suppose some labels $\mathcal{L} = \{l_1, \ldots, l_n\}$ and a distance function $d\colon \mathcal{F} \times \mathcal{L} \to \mathbb{R}$ are given. Then the data dependent distance matrix $D \in \mathbb{R}^{m\times n}$ measuring the fit of labels to the data is defined as

$$D_{ij} := d(f_i, l_j) \quad \text{for} \quad i \in [m], j \in [n]. \tag{3.1}$$

These distances are lifted to the assignment manifold by the lifting map

$$L_i = L_i(W_i, D_i) := \exp_{W_i}(-\tfrac{1}{\rho} D_i), \tag{3.2}$$

where D_i denotes the i-th row of D. These vectors are used to build the similarity matrix $S \in \mathcal{W}$, where each row $S_i = S_i(W, D)$ is the Riemannian mean given by the unique minimizer of

$$\frac{1}{2} \sum_{j \in \mathcal{N}(i)} \omega_{ij} d_{\mathcal{S}}^2\big((p, L_j(W_j, D_j)\big) \tag{3.3}$$

for some chosen weights $\omega_{ij} > 0$ with $\sum_{j \in \mathcal{N}(i)} \omega_{ij} = 1$ for every $i \in \mathcal{V}$. A label assignment is inferred by maximizing the correlation between the current assignment state W and the similarity matrix S and has the form

$$\sup_{W \in \mathcal{W}} E(W) := \langle W, S(W, D) \rangle. \tag{3.4}$$

Adopting the simplifying assumption that averaging changes slowly and approximating the Riemannian mean by the geometric mean, one arrives at the assignment flow formulation (1.1).

Variational Model. Intuitively, the above described assignment flow also contains the two classical components of labeling, using the data for label decisions while locally regularizing these decisions in a spatial neighborhood. The goal for our variational approach is to disentangle these two components in the assignment flow, while keeping the basic building blocks: the overall smooth geometric setting and regularization through geometric local averages.

Using the definition of the distance matrix D from (3.1), we propose the following quality measure $J(W)$ evaluating an assignment matrix $W \in \mathcal{W}$ by

$$J(W) := J_{\text{data}} + \rho J_{\text{reg}} := \langle W, D \rangle + \frac{\rho}{2} \sum_{ij \in \mathcal{E}} \omega_{ij} d_{\mathcal{S}}^2(W_i, W_j) \tag{3.5a}$$

$$= \sum_{i \in \mathcal{V}} \Big(\langle W_i, D_i \rangle + \frac{\rho}{2} \sum_{j \in \mathcal{N}(i)} \omega_{ij} d_{\mathcal{S}}^2(W_i, W_j) \Big), \tag{3.5b}$$

with weights $\omega_{ij} > 0$ and $\sum_{j \in \mathcal{N}(i)} \omega_{ij} = 1$ for every $i \in \mathcal{V}$ and a parameter $\rho > 0$ regulating the amount of regularization.

The purpose of the data term is to choose the best fit to the data, i.e. the smallest distance at every node $i \in \mathcal{V}$. Regularization is induced by comparing the assignments W_i at every node $i \in \mathcal{V}$ with their neighboring assignments W_j with $j \in \mathcal{N}(i)$, similar to [11] and [4]. The main **take-home message of this paper** is: In contrast to the assignment flow, we do not directly compare the assignments W_i to the Riemannian mean itself (cf. (3.4)), but rather use the Riemannian mean defining objective functions (3.3) as a measure for similarity in a spatial neighborhood. This way we avoid the need for an explicit expression of the Riemannian mean and its derivative while at the same time favoring those assignments similar to the Riemannian mean of neighboring assignments.

A label assignment is inferred by minimizing the obj. function: $\inf_{W \in \mathcal{W}} J(W)$. Since $\langle W, D \rangle$ as well as the Riemannian distance are defined and continuous on the whole space Δ^m, so is the objective function J. Yet, this extension is not

smooth due to the square root in the sphere map ψ, which isomorphically embeds $\mathcal{S}$ into the Euclidean sphere. Because of $J\colon \Delta^m \to \mathbb{R}$ being continuous and Δ^m compact, the existence of minimizers is assured and

$$\inf_{W\in\mathcal{W}} J(W) = \min_{W\in\Delta^m} J(W). \tag{3.6}$$

As a consequence, the minimizer W^* might not be an element of the assignment manifold $\mathcal{W} = \operatorname{rint}(\Delta^m)$ anymore.

Parameter influence of ρ. In the following we analyse the set of minimizers for the two extreme cases of the parameter $\rho > 0$ being close to 0 and very large.

Case $\rho \to 0$: For ρ tending to 0, we loose the regularization term in the limit and obtain a minimization problem which separates over the nodes

$$\min_{W\in\Delta^m} \sum_{i\in\mathcal{V}} \langle W_i, D_i\rangle = \sum_{i\in\mathcal{V}} \min_{W_i\in\Delta} \langle W_i, D_i\rangle. \tag{3.7}$$

Assume that every D_i has a unique minimal entry denoted by $D_{im_i} < D_{ij}$ for $j \neq m_i$. Then the unique minimizer W_i^* of $\langle W_i, D_i\rangle$ is given by $W_i^* = e_{m_i} \in \partial\Delta$ for every $i \in \mathcal{V}$. Thus, the objective function J has a unique minimum at a corner point of Δ^m, i.e. W^* is an integral labeling.

Case $\rho \to \infty$: Equivalently characterizing the set of minimizer W^* by

$$\operatorname{argmin}_{W\in\Delta^m} J_{\text{data}}(W)+\rho J_{\text{reg}}(W) = \operatorname{argmin}_{W\in\Delta^m} \tfrac{1}{\rho} J_{\text{data}}(W)+J_{\text{reg}}(W), \tag{3.8}$$

shows, that for ρ tending towards infinity the influence of the data term vanishes in the limit and we are only minimizing with respect to the regularizer, i.e.

$$\operatorname{argmin}_{W\in\Delta^m} \frac{1}{2} \sum_{ij\in\mathcal{E}} \omega_{ij} d_{\mathcal{S}}^2(W_i, W_j). \tag{3.9}$$

In this case the set of minimizers is given by all $W^* \in \Delta^m$ having identical rows, $W_i^* = W_j^*$ for all $i, j \in \mathcal{V}$, showing the existence of interior optima.

Due to the behaviour of the model in these two extreme cases, it is expected that for larger ρ there exist local optima in the interior constituting fully probabilistic assignments while for decreasing ρ closer to 0 local optima tend to lie closer to the corners, i.e. result in integral assignments. Experimentally, this intuition is indeed confirmed (see Experiments below).

Enforcing Integrality and Connection to Graphical Models. In order to enforce (approximate) integral solutions of the model, also called *rounding*, we add an additional cost term punishing large deviations from integral assignments similar to [6]. This can be done by adding the entropy of W defined by

$$H(W) := \sum_{i\in\mathcal{V}} H(W_i) = -\langle W, \log(W)\rangle \in [0, m\log(n)], \tag{3.10}$$

thus obtaining the extended model with integrality enforcing parameter $\alpha > 0$

$$J_\alpha(W) := J(W) + \alpha H(W) = \langle W, D\rangle + \frac{\rho}{2}\sum_{ij\in\mathcal{E}} \omega_{ij} d_{\mathcal{S}}^2(W_i, W_j) + \alpha H(W). \quad (3.11)$$

For $\alpha \to 0$ we obtain the the previous model. For $\alpha \to \infty$ we end up with the following minimization problem

$$\min_{W\in\Delta^m} J(W), \quad \text{s.t.} \quad W_i \in \{0,1\}^n \quad \text{for all} \quad i \in \mathcal{V}. \quad (3.12)$$

Under these constraints, every assignment vector equals a standard basis vector $W_i \in \{e_1, \ldots, e_n\} \subset \mathbb{R}^n$ and the Riemannian distance in the regularizer only takes on two different values

$$d_{\mathcal{S}}^2(W_i, W_j) = \begin{cases} \pi^2 & , \text{for } W_i \neq W_j \\ 0 & , \text{for } W_i = W_j. \end{cases} \quad (3.13)$$

Therefore, this minimization problem has the form of a graphical model with Potts prior (cf. [10, Section 3.3]). It is more appropriate however, to consider our approach as a geometric alternative to the continuous cut variational formulation of the image segmentation problem [5,8]. For an in-depth discussion of evaluating discrete graphical models using the assignment flow we refer to [6].

4 Optimization Approach

Riemannian Gradient Flow. Our optimization strategy is to follow the Riemannian gradient descend flow of J on the manifold $\mathcal{W}$ with a natural unbiased initialization given by the barycenter at every node

$$\dot{W}(t) = -\nabla_{\mathcal{W}} J_\alpha(W(t)), \quad W(0) = \mathbb{1}_m c^\top \quad (4.1)$$

or due to (2.10) equivalently the transformed flow on $\mathcal{T}$

$$\dot{V}(t) = -\nabla J_\alpha(W(t)), \quad V(0) = 0, \quad (4.2)$$

with $W(t) = \exp_C(V(t))$.

Proposition 1. *Setting $\overline{\omega}_{ij} = \frac{1}{2}(\omega_{ij} + \omega_{ji})$, the i-th row of the Riemannian gradient and the Euclidean gradient of J_α at $W \in \mathcal{W}$ are given by*

$$\big(\nabla_{\mathcal{W}} J_\alpha(W)\big)_i = \Pi_{W_i} D_i - \rho \sum_{j\in\mathcal{N}(i)} \overline{\omega}_{ij} \operatorname{Exp}_{W_i}^{-1}(W_j) - \alpha \Pi_{W_i} \log(W_i) \quad (4.3a)$$

$$\big(\nabla J_\alpha(W)\big)_i = P_T D_i - \rho \sum_{j\in\mathcal{N}(i)} \overline{\omega}_{ij} P_T\big[\tfrac{1}{W_i} \operatorname{Exp}_{W_i}^{-1}(W_j)\big] - \alpha P_T \log(W_i). \quad (4.3b)$$

Proof. A standard calculation shows $\nabla\big(J_{\mathrm{data}} + \alpha H\big)(W) = P_T D - \alpha P_T \log(W)$ and therefore $\big(\nabla_{\mathcal{W}}(J_{\mathrm{data}} + \alpha H)(W)\big)_i = \Pi_{W_i} D_i - \alpha \Pi_{W_i} \log(W_i)$ by (2.9) and $\Pi_{W_i} P_T = \Pi_{W_i}$. As for the Riemannian gradient of J_{reg}, we note that the Riemannian gradient of $d^2_{\mathcal{S}}(p,q)$ with respect to p has the form $\nabla_{\mathcal{S},p} d^2_{\mathcal{S}}(p,q) = -2\operatorname{Exp}_p^{-1}(q)$. The Euclidean gradient then follows by applying $\Pi_{W_i}^{-1}$ to (2.9). □

Numerical Integration of the Flow. For simplicity of notation we denote the i-th row of the Euclidean gradient by $\nabla_i J_\alpha(W)$. We follow [9] for discretizing the gradient flow (4.1) on $\mathcal{W}$. This is done by choosing the common *explicit Euler method* for the transformed flow (4.2) on the linear space $\mathcal{T}$, which reads

$$V_i^{(k+1)} = V_i^{(k)} - h^{(k)} \nabla_i J_\alpha(W^{(k)}), \quad V_i^{(0)} = 0 \quad \text{for all} \quad i \in \mathcal{V}, \tag{4.4}$$

where $h^{(k)} > 0$ denotes the step-size. Transforming this update scheme back onto $\mathcal{W}$ by $W_i^{(k)} = \exp_c(V_i^{(k)})$ with initial condition $W_i^{(0)} = \frac{1}{n}\mathbb{1}_n$ and using the fact that $\nabla_i J_\alpha = \nabla_i J(W) - \alpha P_T \log(W_i)$, we obtain a *multiplicative update*

$$W_i^{(k+1)} = \frac{1}{Z_i} W_i^{(k)} e^{-h^{(k)} \nabla_i J_\alpha(W^{(k)})} = \frac{1}{Z_i'} \big(W_i^{(k)}\big)^{1+h^{(k)}\alpha} e^{-h^{(k)} \nabla_i J(W^{(k)})} \tag{4.5}$$

where Z_i and Z_i' are normalizing constants ensuring $\langle W_i^{(k)}, \mathbb{1}\rangle = 1$. This update formula clearly illustrates the influence of the integrality parameter α as some sort of built in rounding mechanism of the flow.

Assignment Normalization. The flow $W(t)$ solving (4.1) evolves on the manifold $\mathcal{W}$ and therefore all entries $W_{ij}(t) > 0$ are positive all the time. If $W_i(t)$ approaches an integral label, then all but one entries approach 0, for $t \to \infty$. However, the multiplicative update (4.5) is only valid on the manifold $\mathcal{W}$. Since there is a difference between mathematical and numerical positivity we adopt the strategy of [2, Sec. 3.3] for ensuring numerical positivity of the discretized flow. The basic idea is to do an ε-*normalization* every time an entry of W_i drops below $\varepsilon > 0$ given by

$$W_i' = W_i + (\varepsilon - \min_j W_{ij})\mathbb{1}_n \quad \text{and} \quad W_i \leftarrow \tfrac{1}{\langle W_i', \mathbb{1}\rangle} W_i'. \tag{4.6}$$

As shown in [3], the ε-normalization has an influence for the discrete flow of [2]. Since the model (3.5) is continuous on Δ^m, applying this normalization strategy only has a negligible effect in our situation for ε close to 0.

5 Experiments

We assess the parameter influence of ρ and α by applying our geometric variational approach to the following image labeling problem. For this, we take a grid graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with neighborhood size $|\mathcal{N}(i)| = 3 \times 3$ for every $i \in \mathcal{V}$

representing the spatial structure of a noisy RGB-image $f\colon \mathcal{V} \to [0,1]^3$, depicted in Fig. 5.2. Eight prototypical colors $\{l_1, \ldots, l_8\} \subset [0,1]^3$ (Fig. 5.2) were used as labels. The distance function $d(f_i, l_j) = \|f_i - l_j\|_1/3$ is used for constructing the distance matrix (3.1) and the weights in the regularizer are set to $\omega_{ij} = 1/|\mathcal{N}(i)|$. For obtaining a fine discretization of the flow, a constant step-size with value $h = 0.1$ is used for numerical integration by applying (4.5) together with $\varepsilon = 10^{-10}$ for the assignment normalization (4.6). As convergence criterion we use the normalized relative change of the objective function J_α, defined by $r(W^{(k)}) := |J_\alpha(W^{(k)}) - J_\alpha(W^{(k-1)})|/|hJ_\alpha(W^{(k-1)})|$. Due to the small (save) step-size and in order to give the dynamics enough time to trade off regularization against rounding in our model, we set the maximum number of iterations to 2000 and stop the algorithm if either $r(W^{(k)})$ drops below a threshold of 10^{-4} or if the maximum number of iterations is reached. The figures provide quantitative illustrations of all aspects of the variational model introduced in Section 3. We refer to the figure captions for a detailed discussion.

6 Conclusion

This work clarifies some mathematically open points in connection with the assignment flow. Due to the way in which we utilized the smooth geometric setting and regularization through geometric local averages, the proposed variational model (3.5) avoids the need for an explicit expression of the Riemannian mean and its derivative. To enforce integral assignments, an extended version of the model was introduced in (3.11), by adding an entropy term. The interaction between regularization and entropy minimization results in an interesting dynamics of the Riemannian gradient flow, illustrated by preliminary experiments. The question of convergence properties and a closer investigation of the interplay between entropy and regularization in the optimization process provides an opportunity for further research.

References

1. Absil P.-A., Mathony R.,Sepulchre R.: *Optimization Algorithms on Matrix Manifolds*. Princeton, N.J. Woodstock: Princeton University Press, 2008.
2. Åström, F., Petra, S., Schmitzer, B., Schnörr, C.: *Image Labeling by Assignment*, J. Math. Imag. Vison, Vol. 58, No. 2, (2017): 211–238.
3. Bergmann, R., Fitschen, J. H., Persch, J., Steidl, G.: *Iterative Multiplicative Filters for Data Labeling*, Int. J, Comp. Vis., Vol. 123, No. 3, (2017): 435–453.
4. Bergmann, R., Tenbrinck, D.: *A Graph Framework for Manifold-Valued Data*, SIAM J. Imag. Sci., Vol. 11, No. 1, (2018): 325–360.
5. Chan, T. F., Esedoglu, S., Nikolova, M.: *Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models*, SIAM J. Appl. Math., Vol. 66, No. 5, (2006): 1632–1648.
6. Hühnerbein, R., Savarino, F., Åström, F. Schnörr, C.: *Image Labeling Based on Graphical Models Using Wasserstein Messages and Geometric Assignment*, SIAM J. Imag. Sci., Vol. 11, No. 2, (2018): 1317–1362.

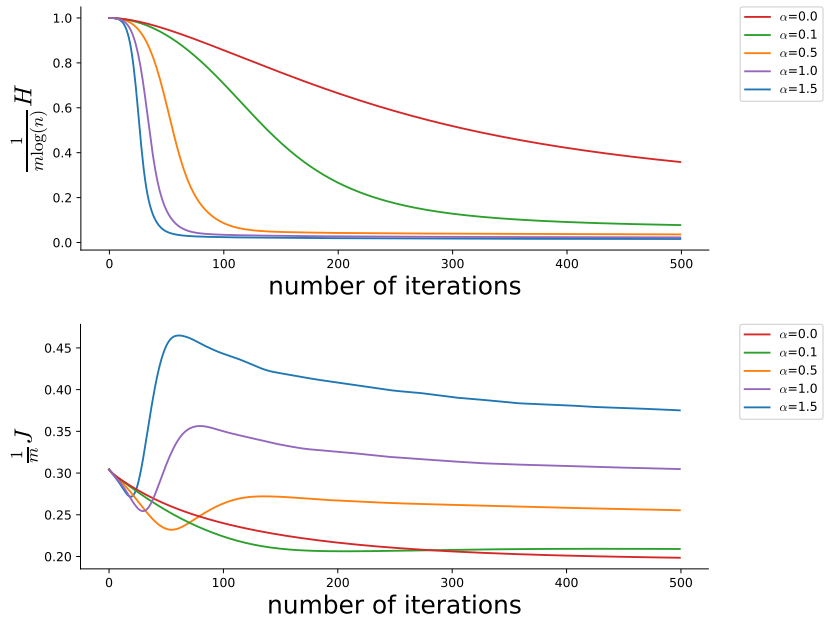

Fig. 5.1: **Convergence and rounding.** The normalized average entropy $\frac{1}{m\log(n)}H$ and average energy $\frac{1}{m}J$ for $\rho = 1.5$ and varying α are shown for the first 500 iterations of the algorithm. TOP: Larger values of α causes a faster decay in entropy which forces the flow to converge more rapidly to integral assignments. BOTTOM: Two phases can be identified depending on the parameter α. In the first phase, the flow moves towards a local optima of J up to the point where the dynamics of the added entropy term takes over in the second phase and pushes the flow towards an integral assignment, leading to higher J values. For smaller values of α, the algorithm spends more time minimizing J while for larger α a faster convergence towards integral solutions is favoured.

7. Kappes, J.H., Andres, B., Hamprecht, F.A., Schnörr, C., Nowozin, S., Batra, D., Kim, S., Kausler, B.X., Kröger, T., Lellmann, J., Komodakis, N., Savchynskyy, B., Rother, C.: *A Comparative Study of Modern Inference Techniques for Structured Discrete Energy Minimization Problems*, IJCV, Vol. 115, No. 2, (2015): 155–184.
8. Lellmann, J., Schnörr, C.: *Continuous Multiclass Labeling Approaches and Algorithms*, SIAM J. Imag. Sci., Vol. 4, No. 4, (2011): 1049–1096.
9. Savarino, F., Hühnerbein, R., Åström, F., Recknagel, J., Schnörr, C.: *Numerical integration of Riemannian gradient flows for image labeling*, Int. Conference on Scale Space and Variational Methods in Computer Vision, Springer, (2017): 361–372
10. Wainwright, M. J., Jordan, M. I.: *Graphical Models, Exponential Families, and Variational Inference*, Found. Trends Mach. Learning, Vol. 1, No. 1–2, (2008): 1–305.
11. Weinmann, A., Demaret, L., Storath, M.: *Total Variation Regularization for Manifold-Valued Data*, SIAM J. Imag. Sci., Vol. 7, No. 4, (2014): 2226–2257.

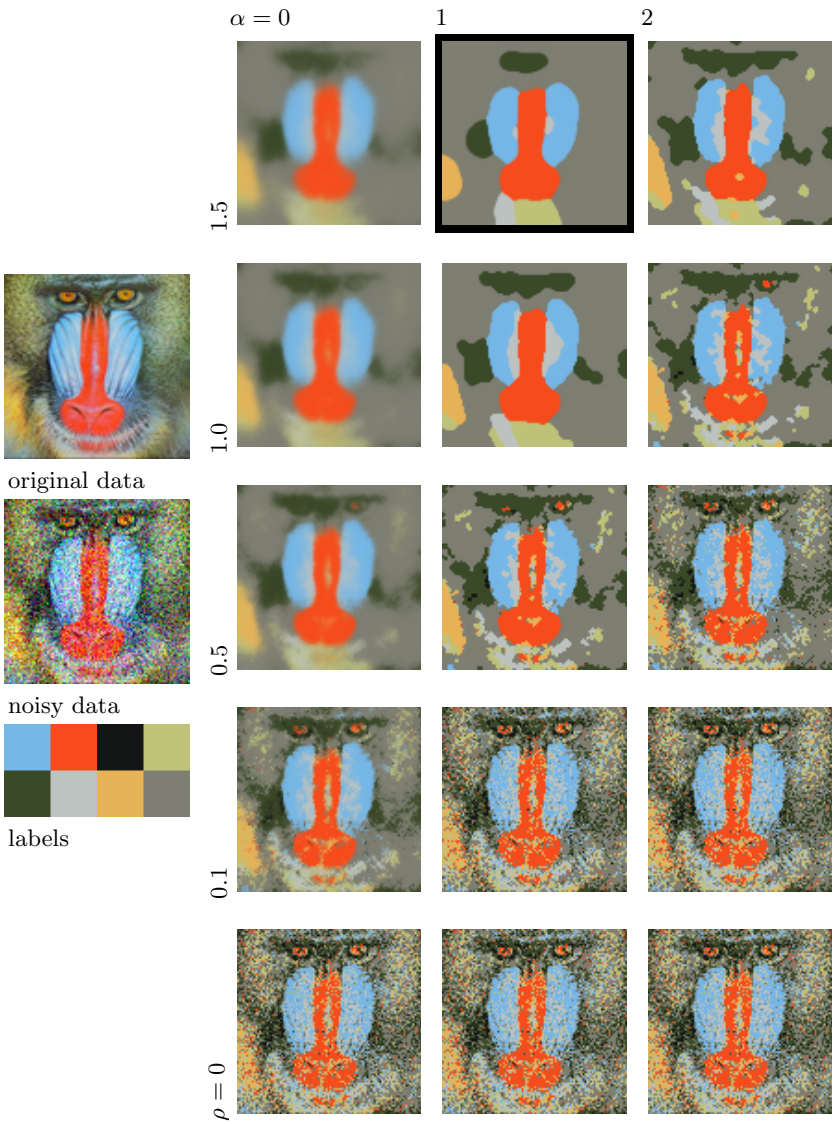

Fig. 5.2: **Parameter influence.** Influence of the parameter ρ controlling the amount of regularization and the parameter α enforcing integrality of the solution. The images depict the expected value of the labels given the assignment W after numerical optimization, i.e. the value of pixel i is given by $\mathbb{E}_{W_i}[\mathcal{L}] = \sum_{j\in[n]} W_{ij} l_j$. The flow corresponding to the image in the top row framed in black reached the maximum number of 2000 iteration without $r(W)$ dropping below 10^{-4}. For $\alpha = 0$, even relatively small values $\rho = 0.1$ and larger ones lead to non integral solutions as indicated by the blurred images in the first column, while for $\rho = 0$ we obtain an integral solution without regularization, as expected from (3.7). The other parameter values for $\alpha > 0$ clearly illustrate the mechanism explained in Fig. 5.1. For smaller α more time is spent to minimize J during the algorithm resulting in more regularized label assignments. Larger α values cause a faster integral decision in an earlier stage of the algorithm leading to less regularized results.

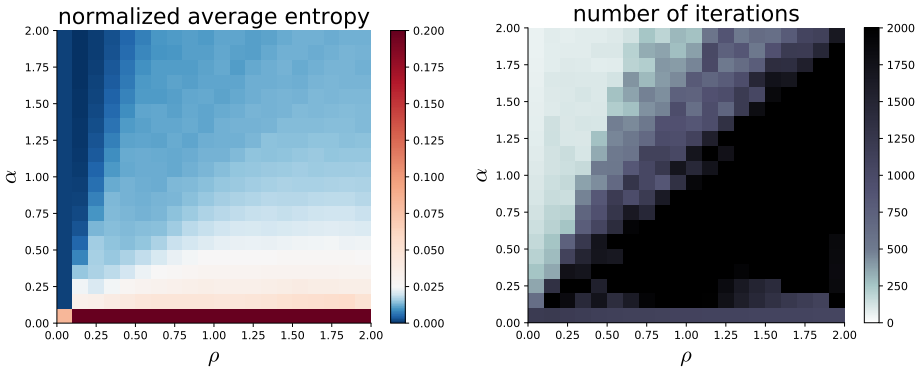

Fig. 5.3: **Interior vs. integral minima.** Evaluation of model (3.11) for all parameter combinations (ρ, α) with values $\rho, \alpha \in \{0.0, 0.1, 0.2, \ldots, 1.9, 2.0\}$. LEFT: The normalized average entropy values $\frac{1}{m \log(n)} H$ after numerical optimization are displayed. As expected from the model, larger α values lead to a decrease in entropy favouring integral solutions as discussed in Fig. 5.1. If ρ increases relative to α then the entropy term weakens and local optima tend towards fractional assignments in the interior, leading to higher entropy values. Two special cases are clearly visible. For $\alpha = 0$ and $\rho > 0$ we are in the regime of (3.5) where local optima tend to lie in the interior causing large entropy values. For $\alpha = \rho = 0$ the optimization problem seperates over the nodes and the flow theoretically converges towards an integral solution as discussed after (3.7) and illustrated in Fig. 5.2 bottom left. However, the numerical integration scheme (4.6) slows down for $\alpha = 0$ as the flow converges towards an integral solution, preventing it from reaching a low entropy state before the termination criterion of the algorithm is fulfilled. RIGHT: The corresponding number of iterations are shown. For $\alpha = 0$ the termination criterion is reached after about 1000 iterations, while for $\rho > \alpha > 0$ the averaging and integrality enforcing effects compensate each other and result in a slowdown of the algorithm, thus reaching the maximum number of 2000 iterations as indicated by the black region. For $\alpha = \rho$ and increasing α the entropy term dominates and accelerates the integration of the flow, leading to faster convergence.