Numerical Integration of Riemannian Gradient Flows for Image Labeling

Fabrizio Savarino$^{(\boxtimes)}$, Ruben Hühnerbein, Freddie Åström, Judit Recknagel, and Christoph Schnörr

Image and Pattern Analysis Group, RTG 1653,
Heidelberg University, Heidelberg, Germany
`{fabrizio.savarino,freddie.astroem}@iwr.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 was proposed [2] by following the Riemannian gradient flow of a given objective function on the so-called assignment manifold. In this paper, we adopt an approach from the literature on uncoupled replicator dynamics and extend it to the geometric labeling flow, that couples the dynamics through Riemannian averaging over spatial neighborhoods. As a result, the gradient flow on the assignment manifold transforms to a flow on a vector space of matrices, such that parallel numerical update schemes can be derived by established numerical integration. A quantitative comparison of various schemes reveals a superior performance of the adaptive scheme originally proposed, regarding both the number of iterations and labeling accuracy.

Keywords: Image labeling · Assignment manifold · Riemannian gradient flow · Replicator equation · Multiplicative updates

1 Introduction

Overview, Motivation. The image labeling problem can be described as assigning to each pixel a single element from a finite set of predefined labels. Usually, this is done by finding optima of a globally defined objective function which evaluates the quality of labelings. In general the problem of computing globally optimal labels is NP-hard. Therefore, various relaxations are used to yield a computationally feasible problem [7].

In [2] a smooth geometric approach is proposed 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 Sect. 2). Their approach is as follows. The assignment manifold $\mathcal{W}$ is turned into a Riemannian manifold by the Fisher Rao

(information) metric (cf. [5]). The i-th element of a given label set $\mathcal{L} = \{l_1, \ldots l_n\}$ is identified with the i-th corner of the standard simplex, i.e. l_i is identified with the i-th standard unit vector $e_i \in \mathbb{R}^n$. For a chosen distance function, the distance information between the image data and predefined labels $\mathcal{L}$ is collected into distance matrix D. This matrix is lifted onto the manifold $\mathcal{W}$ and denoted by L. Riemannian means (also called Karcher means) are used to transform the lifted distance information L into a similarity matrix S, which induces regularizing dependencies between labels assigned to pixels within a spatial neighbourhood. For a more detailed introduction of the assignment filter we refer again to [2].

The quality of an assignment $W \in \mathcal{W}$ is measured by the correlation with the similarity matrix $S(W) \in \mathcal{W}$. Using the standard inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^{m \times n}$, the corresponding objective function is given by

$$J \colon \mathcal{W} \to \mathbb{R}, \quad W \mapsto J(W) := \langle W, S(W) \rangle. \tag{1.1}$$

Finding an optimal assignment corresponds to solving the nonlinear smooth optimization problem $\max_{W \in \mathcal{W}} J(W)$. Their optimization approach is to maximize the objective function by following the *Riemannian gradient ascent flow* on the manifold $\mathcal{W}$,

$$\dot{W}(t) = \nabla_{\mathcal{W}} J(W(t)), \quad W(0) = \frac{1}{n}\mathbb{E}, \tag{1.2}$$

where $\mathbb{E}$ is the matrix containing one in every entry. This constitutes an unbiased initialization in every row i at the barycenter of the respective simplex. Due to the specific choice of the Fisher Rao metric, the gradient flow (1.2) can be rewritten as a *coupled non-linear* system of replicator equations for each row W_i, with $i \in [m] := \{1, \ldots, m\}$, given by

$$\dot{W}_i(t) = W_i(t) \cdot \nabla_i J\big(W(t)\big) - \langle W_i(t), \nabla_i J\big(W(t)\big)\rangle W_i(t), \tag{1.3}$$

where $\nabla_i J(W) = (\frac{\partial}{\partial W_{i1}} J(W), \ldots, \frac{\partial}{\partial W_{in}} J(W))$ is the Euclidean gradient and '$\cdot$' denotes the componentwise multiplication of two vectors.

In [2, Sect. 3] an *explicit Euler method* is used to approximate the integral curve of the gradient flow with the *adaptive step-size* of the i-th row explicitly chosen as

$$h_i^{(k)} = \frac{1}{\langle W_i^{(k)}, \nabla_i J(W^{(k)})\rangle}, \qquad W_i^{(k+1)} = \frac{W_i^{(k)} \cdot \nabla_i J(W^{(k)})}{\langle W_i^{(k)}, \nabla_i J(W^{(k)})\rangle}, \quad i \in [m], \tag{1.4}$$

which results in a multiplicative update scheme for $W_i^{(k)}$. This scheme is then further simplified by approximating the Euclidean gradient with the similarity matrix $\nabla_i J(W) \approx S_i(W)$. An obvious advantage of this update formula is that it is easy to implement and computationally cheap. As demonstrated in [2] this numerical scheme achieved good performance on some academical examples, despite its simplicity.

On the other hand, this numerical scheme *only works for this particular setting*: choice of the Fisher Rao metric, explicit Euler updates with a specific

step-size rule, the particular objective function J and the gradient approximation $\nabla_i J(W) \approx S_i(W)$. As a result, if any of these ingredients is changed, the proposed scheme of [2] is not applicable anymore.

Contribution. In this paper we propose a more principled approach in terms of a top down numerical framework. To this end, we generalize the *transformation* of the uncoupled replicator equation from [3] to gradient flows *of arbitrary objective functions on any Riemannian manifold*. This transformation is then applied to the assignment manifold. The main idea is to transform the gradient flow from the assignment manifold $\mathcal{W}$ onto a vector subspace $T^m \subset \mathbb{R}^{m \times n}$ of $m \times n$ matrices, using a diffeomorphism $\exp_C : T^m \to \mathcal{W}$ (see Fig. 1). This transformation ensures that the corresponding numerical solution of the gradient flow, computed on T^m, evolves on the assignment manifold $\mathcal{W}$. In this framework, the numerical algorithm can be flexibly chosen and adapted for any objective function and any Riemannian metric on the assignment manifold.

As a second contribution, we use our framework for generating and comparing various multiplicative update schemes, their efficiency regarding the number of iterations, and how they effect the labeling accuracy.

(a)   (b)
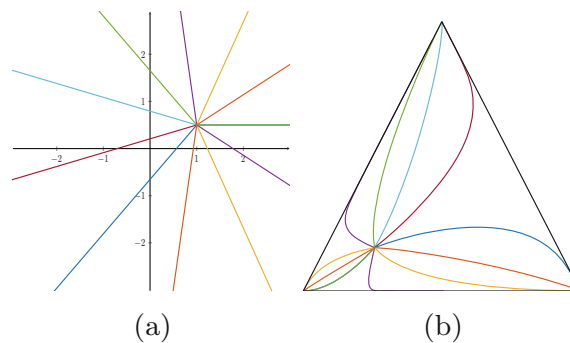
Fig. 1. Illustration of the transformation map. For $m = 1$, rays in T originating from a point $x \in T$ (a) are mapped by $\exp_C$ to curves on the simplex (b) starting at $p = \exp_C(x)$. This curves are first-order approximation of the geodesics at p (see Sect. 2).

2   Preliminaries: The Assignment Manifold

We briefly introduce the notation and geometric setting of the assignment manifold from [2]. Let $m, n \in \mathbb{N}_{>0}$. The standard inner product on $\mathbb{R}^n$ as well as on $\mathbb{R}^{m \times n}$ is denoted by $\langle \cdot, \cdot \rangle$ and $\mathbb{1} = (1, 1, \ldots, 1)^\top \in \mathbb{R}^n$. Let $\Delta_n = \{p \in \mathbb{R}^n : p_i \geq 0 \text{ for } i = 1, \ldots, n, \langle p, \mathbb{1} \rangle = 1\}$ be the probability simplex of dimension $n - 1$. The relative interior of the simplex

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

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

$$T_p \mathcal{S} = \{v \in \mathbb{R}^n : \langle v, \mathbb{1} \rangle = 0\} =: T \subset \mathbb{R}^n. \qquad (2.2)$$

There is an orthogonal decomposition $\mathbb{R}^n = T \oplus \mathbb{R}\mathbb{1}$ with respect to $\langle \cdot, \cdot \rangle$ together with the linear projection map onto T given by

$$\pi \colon \mathbb{R}^n \to T, \quad x \mapsto \pi[x] := \left(I - \tfrac{1}{n}\mathbb{1}\mathbb{1}^\top\right)x. \qquad (2.3)$$

$\mathcal{S}$ equipped with the Fisher Rao metric becomes a Riemannian manifold. The Fisher Rao metric at $p \in \mathcal{S}$ is given by

$$g_p^{\mathcal{S}} \colon T \times T \to \mathbb{R}, \qquad g_p^{\mathcal{S}}(u, v) = \langle u, \mathrm{Diag}(\tfrac{1}{p})v \rangle, \qquad u, v \in T. \qquad (2.4)$$

Since there is a global chart for $\mathcal{S}$ and constant tangent space T, the tangent bundle is trivial $T\mathcal{S} \cong \mathcal{S} \times T$. In order to assure closed form solutions to Riemannian means, the exponential map of $\mathcal{S}$ is approximated in [2] by the so-called *lifting map*

$$\exp \colon T\mathcal{S} = \mathcal{S} \times T \to \mathcal{S}, \quad (p, u) \mapsto \exp_p(u) = \frac{pe^u}{\langle p, e^u \rangle}, \qquad (2.5)$$

where e^u is the componentwise exponential. Although, $\exp_p \colon T \longrightarrow \mathcal{S}$ is a diffeomorphism for every $p \in \mathcal{S}$, it is not the exponential map of the Riemannian manifold. However, according to [2, Proposition 3.1] it provides a first-order approximation of the geodesics at $p \in \mathcal{S}$.

The *assignment manifold* $\mathcal{W}$ is defined to be the product manifold given by (2.6). Using $T_p\mathcal{S} = T$ and the usual identification of the tangent space of a product manifold, the tangent space $T_W\mathcal{W}$ of $\mathcal{W}$ at $W \in \mathcal{W}$ is

$$\mathcal{W} := \prod_{i=1}^m \mathcal{S}, \qquad T_W\mathcal{W} = \prod_{i=1}^m T_{W_i}\mathcal{S} = \prod_{i=1}^m T =: T^m. \qquad (2.6)$$

An element $W = (W_1, \ldots, W_m)^\top \in \mathcal{W} \subset \mathbb{R}^{m \times n}$ is a $m \times n$ matrix with $W\mathbb{1} = \mathbb{1}$, where the i-th row W_i is an element of $\mathcal{S}$. Similarly, $V = (V_1, \ldots, V_m)^\top \in T^m \subset \mathbb{R}^{m \times n}$ is a matrix with $V\mathbb{1} = 0$, where every i-th row V_i is an element of T. In the following this identification of $\mathcal{W}$ and T^m as subsets of $\mathbb{R}^{m \times n}$ is used.

There is again an orthogonal decomposition $\mathbb{R}^{m \times n} = T^m \oplus \{\lambda^\top \mathbb{E} : \lambda \in \mathbb{R}^m\}$ with respect to $\langle \cdot, \cdot \rangle$, where $\mathbb{E} \in \mathbb{R}^{m \times n}$ is the matrix with $\mathbb{E}_{ij} = 1$ for all entries. The orthogonal projection map onto T^m is given by

$$\Pi \colon \mathbb{R}^{m \times n} \to T^m, \quad X \mapsto \Pi[X] := (\pi[X_1], \ldots, \pi[X_m])^\top, \qquad (2.7)$$

with π due to (2.3), and where X_i denotes the i-th row of the matrix $X \in \mathbb{R}^{m \times n}$.

The Fisher Rao metric on every component $\mathcal{S}$ induces a product metric and turns $\mathcal{W}$ into a Riemannian manifold. At $W \in \mathcal{W}$ and for $U, V \in T^m$, this Riemannian metric $g_W^{\mathcal{W}}$ is given by

$$g_W^{\mathcal{W}}(U, V) = \sum_{k=1}^m g_{W_k}^{\mathcal{S}}(U_k, V_k) = \sum_{k=1}^m \langle U_k, \mathrm{Diag}(\tfrac{1}{W_k})V_k \rangle,$$

where W_k, V_k and U_k again denote the k-th row of the matrices W, V and U. The mapping (2.5) naturally extends to $\mathcal{W}$ by (we reuse the symbol exp)

$$\exp \colon \mathcal{W} \times T^m \to \mathcal{W}, \quad (W, V) \mapsto \exp_W(V) = \left(\exp_{W_1}(V_1), \ldots, \exp_{W_m} V_m\right).$$

The map $\exp_W \colon T^m \to \mathcal{W}$ is a diffeomorphism for every $W \in \mathcal{W}$.

3   Transformation of the Gradient Flow

3.1   General Approach: Isometric Manifolds

Before we present the result of this section, we briefly review some basic concepts of differential geometry, based on [1,8].

Let (M, g^M) be a Riemannian manifold and $f : M \to \mathbb{R}$ a smooth function. Using the standard identification $T_r\mathbb{R} = \mathbb{R}$ for $r \in \mathbb{R}$, the Riemannian gradient $\nabla_M f(x)$ of f at $x \in M$ can be defined as the unique element of $T_x M$ satisfying

$$g_x^M(\nabla_M f(x), v) = Df(x)[v], \quad \forall v \in T_x M, \tag{3.1}$$

where $Df(x) : T_x M \to T_{f(x)}\mathbb{R} = \mathbb{R}$ is the differential of f.

Let (N, g^N) be another Riemannian manifold. The *pullback* of the Riemannian metric g^N via a smooth map $F : M \to N$ at $p \in M$ is denoted as $F^* g_p^N : T_p M \times T_p M \to \mathbb{R}$ and defined by

$$\left(F^* g^N\right)_p(v, u) = g_{F(p)}^N\big(DF(p)[v], DF(p)[u]\big), \quad u, v \in T_p M. \tag{3.2}$$

In general, $F^* g^N$ need not be positive definite and hence may not be a Riemannian metric. If $F : M \to N$ is an immersion, however, then $g^M := F^* g^N$ is a Riemannian metric.

An *isometry* is a diffeomorphism $F : (M, g^M) \to (N, g^N)$ between Riemannian manifolds with $F^* g^N = g^M$. Using standard arguments from differential geometry we have the following.

Lemma 1. *Let $F : (M, g^M) \to (N, g^N)$ be an isometry between two Riemannian manifolds and $f : N \to \mathbb{R}$ as well as $\bar{f} : M \to \mathbb{R}$ be smooth real valued functions with $\bar{f} = f \circ F$, i.e. the diagram commutes. Then the Riemannian gradients of f and $\bar{f}$ are related for every $x \in M$ by $\nabla_N f(F(x)) = DF(x)[\nabla_M \bar{f}(x)].$*
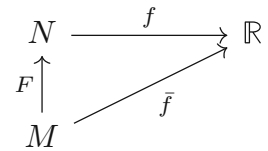
An immediate consequence of the previous lemma is the following.

Proposition 1. *Suppose (M, g^M) and (N, g^N) are isometric Riemannian manifolds with isometry $F : M \to N$. Let $f : N \to \mathbb{R}$ be a smooth function and set $\bar{f} = f \circ F : M \to \mathbb{R}$. Furthermore, let $C \subset \mathbb{R}$ be an open interval with $0 \in C$. Then $\gamma : C \to N$ with $\gamma(0) = y_0$ solves (3.3)(a) if and only if the curve $\eta : C \to M$ with $\gamma = F \circ \eta$ and $\eta(0) = F^{-1}(y_0)$ solves (3.3)(b).*

$$\text{(a)} \quad \dot{\gamma}(t) = \nabla_N f(\gamma(t)) \qquad \text{(b)} \quad \dot{\eta}(t) = \nabla_M \bar{f}(\eta(t)). \tag{3.3}$$

3.2   Application to the Assignment Manifold

Transformation for the Assignment Manifold. The idea of transforming a gradient flow on $\mathcal{W}$ onto a vector space is as follows. Set $C := \frac{1}{n}\mathbb{E} \in \mathcal{W}$,

i.e. every row $C_i = \frac{1}{n}\mathbb{1} \in \mathcal{S}$ is the barycenter of the simplex. The tangent space $T_C\mathcal{W} = T^m \subset \mathbb{R}^{m\times n}$ is a vector subspace and itself a smooth manifold. Using the diffeomorphism $\exp_C\colon T^m \to \mathcal{W}$, we turn T^m into a Riemannian manifold through the pullback metric $g^{T^m} := \exp_C^* g^{\mathcal{W}}$. By construction, $\exp_C$ is an isometry between (T^m, g^{T^m}) and $(\mathcal{W}, g^{\mathcal{W}})$.

Suppose now $f\colon \mathcal{W} \to \mathbb{R}$ is a smooth function and set $\bar{f} := f \circ \exp_C\colon T^m \to \mathbb{R}$. Let $\nabla_{\mathcal{W}} f$ and $\nabla_{T^m}\bar{f}$ denote the Riemannian gradient of $(\mathcal{W}, g^{\mathcal{W}})$ and (T^m, g^{T^m}) respectively. Then we can transform the gradient flow of f due to Proposition 1.

Corollary 1. *Given the setting above, the curve $t \mapsto W(t) \in \mathcal{W}$ with $W(0) = C$ solves the gradient flow* (3.4)(a) *if and only if the curve $t \mapsto V(t) \in T^m$ with $V(0) = 0$ and $W(t) = \exp_C\big(V(t)\big)$ solves the gradient flow* (3.4)(b).

$$(a) \quad \dot{W}(t) = \nabla_{\mathcal{W}} f(W(t)) \qquad\qquad (b) \quad \dot{V}(t) = \nabla_{T^m}\bar{f}\big(V(t)\big). \qquad (3.4)$$

Remark 1. The above construction works in general. One can choose any Riemannian metric on $\mathcal{S}$, define the induced product metric on $\mathcal{W}$ and then turn $\exp_C$ into an isometry by using the pullback metric on T^m.

Corollary 1 enables the transformation of the Riemannian gradient flow (1.2), which has the form (3.4)(a), to a flow of the form (3.4)(b) on the vector space $T^m \subset \mathbb{R}^{m\times n}$, to which standard numerical methods can be applied. Furthermore, this formulation ensures that the corresponding flow in terms of $W(t)$ always stays on the assignment manifold, and that there is no need for projecting onto the simplex after a numerical integration step.

Calculating the Riemannian Gradient. In order to compute the gradient flow $\dot{V}(t) = \nabla_{T^m}\bar{f}$ of a pullback function $\bar{f} = f \circ \exp_C$ with respect to the pullback metric $g^{T^m} = \exp_C^* g^{\mathcal{W}}$, we express the Riemannian gradient $\nabla_{T^m}\bar{f}$ on (T^m, g^{T^m}) in terms of the Euclidean gradients $\nabla\bar{f}$ and ∇f with respect to the induced Euclidean structure of $T^m, \mathcal{W} \subset \mathbb{R}^{m\times n}$. To this end, we express first the pullback metric g^{T^m} in terms of the canonical inner product $\langle \cdot, \cdot \rangle$ on $T^m \subset \mathbb{R}^{m\times n}$.

Lemma 2. *Let $U, V \in T^m$. For any $X \in T^m$, the pullback metric g^{T^m} on T^m is given by $g_X^{T^m}(U, V) = \langle U, D\exp_C(X)[V]\rangle$.*

Proof. Set $W = \exp_C(X)$. It follows from the proof of [2, Proposition 3.1] that the differential of the lifting map $\exp_{C_i}\colon T \to \mathcal{S}$ onto the simplex is given by $D\exp_{C_i}(X_i) = \mathrm{Diag}(W_i) - W_i W_i^\top$ with $W_i = \exp_{C_i}(X_i)$ for all $i = 1, \ldots, m$. A short calculation using this explicit formula and the orthogonal projection $\pi\colon \mathbb{R}^n \to T$ given by (2.3) shows $\pi \circ \mathrm{Diag}(\frac{1}{W_i}) \circ D\exp_{C_i}(X_i)[V_i] = V_i$. Insertion into the definition of the pullback metric g^{T^m} proves the statement. $\qquad\square$

Lemma 3. *The Euclidean gradients of $\bar{f} = f \circ \exp_C$ and f at $X \in T^m$ are related by $\nabla\bar{f}(X) = D\exp_C(X)\big[\nabla f\big(\exp_C(X)\big)\big]$.*

Proof. With $W = \exp_C(X)$, we have for all $V \in T^m$

$$\langle \nabla \bar{f}(X), V \rangle = D\bar{f}(X)[V] = Df(W)\big[D\exp_C(X)[V]\big]$$
$$= \langle \nabla f(W), D\exp_C(X)[V] \rangle = \langle D\exp_C(X)[\nabla f(W)], V \rangle,$$

where the last equality holds due to the symmetry of $D\exp_C(X)$ with respect to the Euclidean metric on $\mathbb{R}^{m \times n}$. Thus, $\nabla \bar{f}(X) = D\exp_C(X)[\nabla f(W)]$.   □

Lemma 4. *Given the setting above, the Riemannian gradient of $\bar{f} : T^m \to \mathbb{R}$ with respect to g^{T^m} is given by $\nabla_{T^m}\bar{f}(X) = \nabla f\big(\exp_C(X)\big)$.*

Proof. Let $X \in T^m$. Due to the formula of Lemma 2, we have

$$D\bar{f}(X)[V] = g_X^{T^m}\big(\nabla_{T^m}\bar{f}(X), V\big) = \langle D\exp_C(X)[\nabla_{T^m}\bar{f}(X)], V \rangle, \qquad (3.5)$$

for all $V \in T^m$. Using the induced inner product on $T^m \subset \mathbb{R}^{m \times n}$ yields the Euclidean gradient $\nabla \bar{f}(X)$ with $D\bar{f}(X)[V] = \langle \nabla \bar{f}(X), V \rangle$ for all $V \in T^m$. Combining these two equations gives $\langle \nabla \bar{f}(X), V \rangle = \langle D\exp_C(X)[\nabla_{T^m}\bar{f}(X)], V \rangle$ for all $V \in T^m$ and therefore $D\exp_C(X)[\nabla_{T^m}\bar{f}(X)] = \nabla \bar{f}(X)$. By Lemma 3

$$D\exp_C(X)[\nabla_{T^m}\bar{f}(X)] = \nabla \bar{f}(X) = D\exp_C(X)\big[\nabla f\big(\exp_C(X)\big)\big]$$

and the statement follows, since $D\exp_C(X)$ is a vector space isomorphism.   □

Now we return to the objective function (1.1) from [2], given by $J(W) = \langle W, S(W) \rangle$ for $W \in \mathcal{W}$. We are prepared to use the transformed gradient flow for solving the maximization problem $\max_{W \in \mathcal{W}} J(W)$. Applying Corollary 1 to the gradient flow (1.2) on $\mathcal{W}$ yields the flow $\dot{V} = \nabla_{T^m}\bar{J}\big(V(t)\big)$ on the tangent space, with $\bar{J} = J \circ \exp_C$. Due to Lemma 4, we further have $\dot{V}(t) = \nabla J\big(\exp_C(V(t))\big)$. It remains to compute the Euclidean gradient $\nabla J(W) \in T^m$.

Lemma 5. *The Euclidean gradient of $J : \mathcal{W} \to \mathbb{R}$ with $J(W) = \langle W, S(W) \rangle$ is given by*

$$\nabla J(W) = \Pi\big[S(W)\big] + DS(W)^* \circ \Pi\big[W\big],$$

where $DS(W)^*$ is the adjoint linear map of $DS(W) : T^m \to T^m$ with respect to the Euclidean inner product $\langle \cdot, \cdot \rangle$ on T^m, and $\Pi : \mathbb{R}^{m \times n} \to T^m$ is the orthogonal projection given by (2.7).

Proof. A smooth curve $\gamma : (-\epsilon, \epsilon) \to \mathcal{W}$ with $\gamma(0) = W, \dot{\gamma}(0) = V \in T^m$ gives

$$DJ(W)[V] = \frac{d}{dt}J\big(\gamma(t)\big)\Big|_{t=0} = \langle \dot{\gamma}(t), S\big(\gamma(t)\big) \rangle\Big|_{t=0} + \langle \gamma(t), \frac{d}{dt}S\big(\gamma(t)\big) \rangle\Big|_{t=0}$$
$$= \langle V, S(W) \rangle + \langle W, DS(W)[V] \rangle.$$

Using the projection $\Pi : \mathbb{R}^{m \times n} \to T^m$, we decompose W and $S(W)$ into $W = \Pi[W] + U_W$ and $S(W) = \Pi[S(W)] + U_S$, with $U_W, U_S \in (T^m)^\perp$. Due to $V, DS(W)[W] \in T^m$ these orthogonal decompositions give

$$\langle V, S(W) \rangle + \langle W, DS(W)[V] \rangle = \langle V, \Pi[S(W)] \rangle + \langle \Pi[W], DS(W)[V] \rangle$$
$$= \langle V, \Pi[S(W)] \rangle + \langle DS(W)^* \circ \Pi[W], V \rangle,$$

where $DS(W)^*$ is the adjoint linear map of $DS(W)$. Thus, we have $DJ(W)[V] = \langle \Pi[S(W)] + DS(W)^* \circ \Pi[W], V \rangle, \forall V \in T^m$, which proves the claim.   □

Summing up, we obtain the explicit form of the flow (3.4)(b) in the specific case $\bar{f} = \bar{J} = J \circ \exp_C$.

Theorem 1. *Solving the gradient flow* (1.2) *from* [2] *is equivalent to* $W(t) = \exp_C(V(t))$, *where* $V(t)$ *solves*

$$\dot{V}(t) = \nabla J\big(W(t)\big) = \Pi\big[S(W)\big] + DS(W)^* \circ \Pi\big[W\big], \qquad V(0) = 0. \qquad (3.6)$$

4   Numerical Approach

Multiplicative Update Formulae. The reformulation of the gradient flow on the tangent space enables the application of a broad range of numerical schemes. We discretize the transformed flow using explicit Runge–Kutta methods [6, Chap. II.2]. A general iteration step for each row $i = 1, \ldots, m$ reads

$$V_i^{(k+1)} = V_i^{(k)} + h^{(k)} P_i^{(k)}, \quad W_i^{(k+1)} = \exp_{C_i}\big(V^{(k+1)}\big) = \frac{W_i^{(k)} e^{h^{(k)} P_i^{(k)}}}{\langle W_i^{(k)}, e^{h^{(k)} P_i^{(k)}} \rangle}, \qquad (4.1)$$

where $h^{(k)} \in \mathbb{R}_{>0}$ denotes the step-size and $P_i^{(k)}$ the direction in the k-th iteration. We point out the similarity to the multiplicative updates (1.4) and the ability to modify them, through the choice of a numerical integration method represented by $P_i^{(k)}$ of (4.1).

Assignment Normalization. Let $W(t) \in \mathcal{W}$ be a smooth curve solving the gradient flow (1.2). Based on [2, Conjecture 3.1], every row $W_i(t) \in \mathcal{S}$ of this solution curve is expected to approach some vertex of the simplex for $t \to \infty$. As a consequence, all but one entry of $W_i(t)$ approach 0 as $t \to \infty$. However, the numerical computations also have to evolve on $\mathcal{W}$, i.e. all entries of $W(t)$ have to be positive all the time. Since, there is a difference between mathematical and numerical positivity, we adopt the strategy in [2] to avoid these numerical problems. This is done by restricting the discrete flow of every row $W_i(t)$ onto the ε -simplex $\Delta_\varepsilon := \{p \in \Delta_n : p_i \geq \varepsilon \text{ for } i = 1, \ldots, n\}$ through a normalization step (cf. [2, Sect. 3.3.1]), after each iteration.

Let $V(t) \in T^m$ be the solution of the transformed gradient flow and $W(t) = \exp_C\big(V(t)\big)$. The convergence of each $W_i(t) \in \mathcal{S}$, $i \in [m]$ to some vertex of the simplex translates to the convergence of each $V_i(t) \in T$, $i \in [m]$ to infinity in a certain direction, as depicted in Fig. 1. In order to normalize $W(t)$, we restrict each $V_i(t)$ to the closed ball $B_R(0) \subset T^m$ of radius $R > 0$ centered at 0. This can be seen as a smooth approximation of the ε-simplex by choosing the radius R such that the image of the sphere $\partial B_R(0)$ under $\exp_C$ intersects Δ_ε at the vertices (Fig. 2c).

As shown in [4], the normalization in [2] dramatically influences the discrete flow. Therefore, it is expected that the normalization on the tangent space influences the discrete flow as well. In this paper, however, we investigate the flow up to the smooth ε boundary for $\varepsilon = 10^{-10}$ (Fig. 2c) and leave a numerical analysis of the flow *on* this boundary for follow-up work.
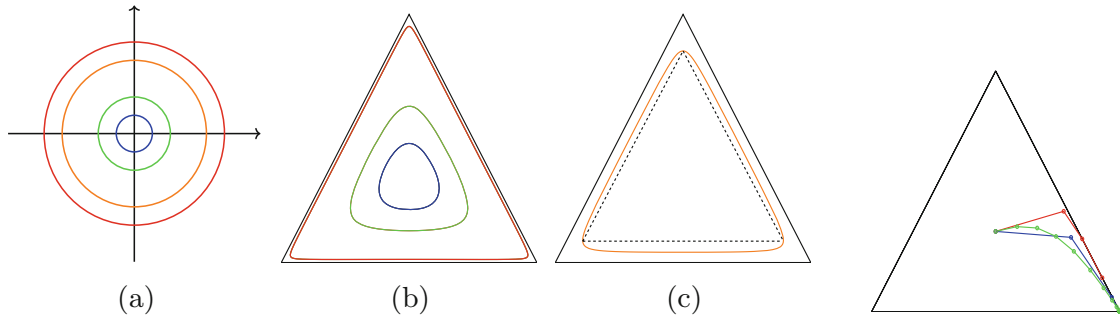
Fig. 2. Assignment Normalization on the Tangent Space. Illustration of how balls centered at 0 on the tangent space T (a) are mapped by each component of $\exp_C \colon T^m \to \mathcal{W}$ onto the simplex (b) in order to smoothly approximate the ε-simplex Δ_ε (c dashed line).
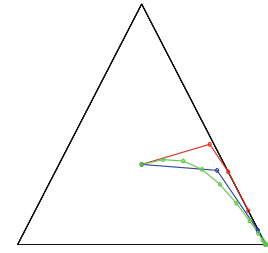
Fig. 3. Typical stepsizes. Manifold approach: red line. Tangent space: green line step length 1, blue line step length 5. (Color figure online)

5 Experiments

In this section we investigate the influence of the discretization method and the specific choice of the step-size in [2], by comparing it to our approach using a more accurate discretization. In [2], an adaptive step-size for the i-th row $W_i^{(k)}$ in the k-th iteration is explicitly set to $h_i^{(k)}$ given by (1.4) in order to arrive at the multiplicative update scheme (1.4) for numerically integrating the gradient flow (1.3). At first glance, this choice seems rather non-intuitive: During the initial phase of the iteration, when W_i and S_i are *uncorrelated*, then $\langle W_i^{(k)}, S_i(W^{(k)}) \rangle \approx 0$ and the step-sizes are *large*, whereas if they are *correlated*, then the step-sizes are *small* and slow down the convergence of the algorithm.

   Our focus in this paper is on the *initial phase of the iteration where by far the most assignments of labels emerge*: Do the aggressive step-sizes (1.4) affect the quality of the resulting labeling?

Set-Up and Implementation. For this assessment, we chose an academical labeling scenario depicted by Fig. 4. The color image comprised 256 colors which also served as labels $\mathcal{C} = \{c_1, \ldots, c_{256}\}$. By shuffling colors at randomly chosen pixel locations we created a noisy version as input for the different methods. The labeling task is to recover the ground truth image. The success of this task is measured by the percentage of correctly labeled pixels.

   In order to compare impartially the discretization methods we also used the simplified gradient on the tangent space $\nabla J(W) \approx \Pi[S(W)]$ (cf. [2, Sect. 3.3.3]). The main intuition behind this simplification is that the similarity matrix $S(W)$ is the result of averaging over spatial neighborhoods and therefore is expected to change slowly, i.e. $DS(W) \approx 0$. Due to this assumption and in view of Theorem 1 the transformed flow simplifies to

$$\dot{V}(t) = \Pi\big[S\big(W(t)\big)\big]. \tag{5.1}$$

Likewise, the similarity matrix $S(W)$ is computationally efficiently approximated by the normalized geometric mean according to [2, Lemma 3.3].

For our experiments we used $\varepsilon = 10^{-10}$ for the ε-simplex normalization, which corresponds to a radius $R \approx 23$ of the closed ball in the tangent space (Fig. 2c). To avoid the aforementioned influence of the *discrete* flows caused by the normalization (cf. Sect. 4), we only compared the solutions after numerical integration *up to the first time where the flow meets the ε -simplices and normalization occurs* (cf. Fig. 2). Due to this termination criterion not all the rows $W_i^{(k)} \in \mathcal{S}$ may have converged to a vertex of the respective simplex.

After termination of every discretization method, we do the following to obtain an unique labeling as output. At every pixel i, we choose label $c_{k_i} \in \mathcal{C}$ with k_i the column index of the maximum entry of $W_i = (W_{i1}, \ldots, W_{i256})$.

For the integration of the gradient flow (5.1) on the tangent space, we considered the common *explicit Euler method*, which reads

$$V_i^{(k+1)} = V_i^{(k)} + h^{(k)}\Pi\big[S\big(W^{(k)}\big)\big], \tag{5.2}$$

and *Heun's method*, which reads

$$\begin{aligned}\tilde{V}_i^{(k+1)} &= V_i^{(k)} + h^{(k)}\Pi\big[S\big(W^{(k)}\big)\big], \\ V_i^{(k+1)} &= V_i^{(k)} + \frac{h^{(k)}}{2}\big(\Pi\big[S\big(W^{(k)}\big)\big] + \Pi\big[S\big(\tilde{W}^{(k+1)}\big)\big]\big).\end{aligned} \tag{5.3}$$

In both cases we us the initial value $V^{(0)} = 0$. For more details about these methods, we refer e.g. to [6, Chap. II.2].

Results. Figure 4 and Table 1 summarize our quantitative findings. The approach [2] ('manifold approach') is compared to the two schemes (5.2) and (5.3) ('tangent space approach') using three different stepsizes for each, and three different scales for spatial regularization (neighbourhood size $|\mathcal{N}_\varepsilon|$ for geometric averaging). During all experiments the selectivity parameter ρ ([2, Sect. 3.1]) for scaling the distance matrix is chosen to be constant. Observations:

(i) Despite early termination after first-time hitting the ε-simplex boundary, $\geq 93\%$ *correct* decisions are made by *all* methods, for a reasonable strength of regularization ($\geq 5 \times 5$ neighbourhoods). The performance is slightly inferior for weak regularization (3×3 neighbourhood), due to the influence of noise. It also deteriorates for larger spatial scales, because then signal structure is regarded as noise, too (compare the slightly decreasing performance of 7×7 vs. 5×5).

(ii) Although the manifold approach takes the *minimal* number of updates (listed as numbers in brackets) due to the aggressive adaptive stepsizes (1.4), it performed *best*!

Observation (ii) is our major – somewhat surprising – finding: A more careful numerical evolution and integration of the Riemannian gradient flow does *not*

$|\mathcal{N}_\varepsilon| = 3 \times 3$ $|\mathcal{N}_\varepsilon| = 5 \times 5$ $|\mathcal{N}_\varepsilon| = 7 \times 7$

Original data

Noisy data
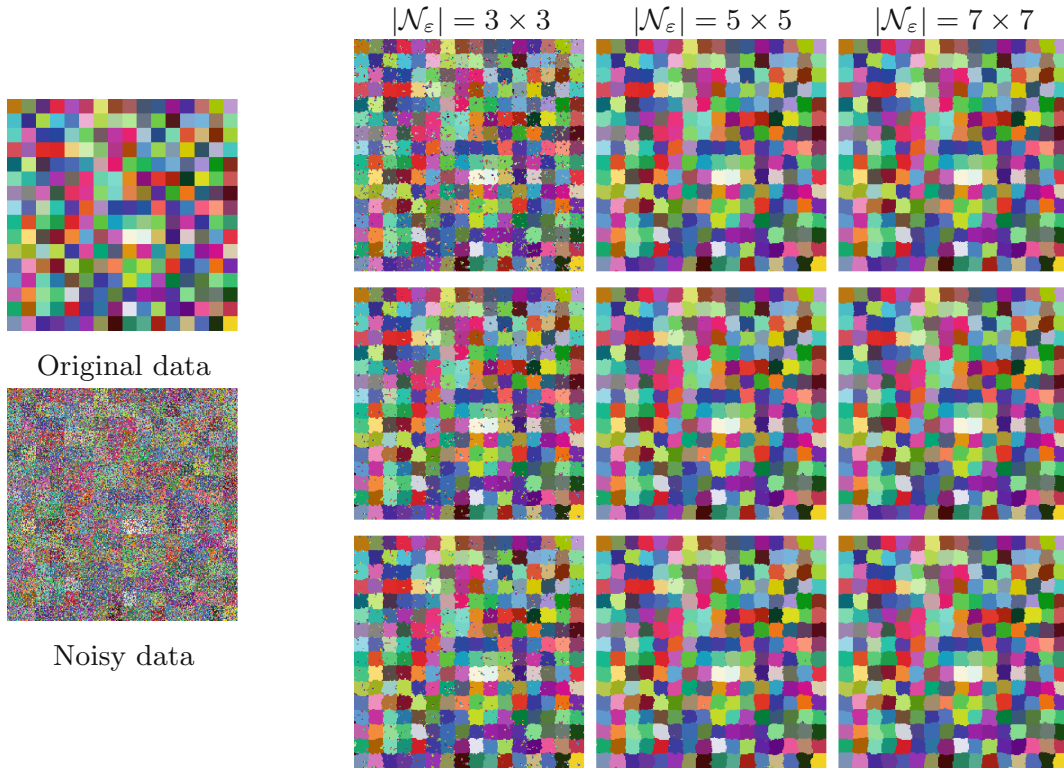
Fig. 4. Step-size influence on labeling. First row: method from [2]. Second row: Explicit Euler method on the tangent space with constant step-size 1. Third row Heun's method on the tangent space with step-size 1. All experiments where done with $\rho = 0.1$. (Color figure online)

Table 1. Labeling performance of geometric flows. The table displays for each integration method the percentages of correctly labeled pixels and the number of iterations in parentheses until the first normalization occurs. h_{adapt} denotes the adaptive step-sizes of [2] given by (1.4). We compare h_{adapt} with constant step-sizes on the tangent space with length $h^1_{\mathrm{const}} = 1$, $h^5_{\mathrm{const}} = 5$ and $h^{10}_{\mathrm{const}} = 10$. For an interpretation of the parameters ρ and $|\mathcal{N}_\varepsilon|$, we refer to [2, Sect. 3.1]. See Sect. 5, paragraph 'Results', for a discussion of the table.

$\rho = 0.1$		Manifold approach	Tangent space approach					
			Explicit Euler			Heun's method		
Step-size		h_{adapt}	h^1_{const}	h^5_{const}	h^{10}_{const}	h^1_{const}	h^5_{const}	h^{10}_{const}
Neighborhood size $|\mathcal{N}_\varepsilon|$	3×3	87,7	90,6	89,4	89,0	91,1	91,0	91,3
		(2)	(24)	(5)	(3)	(24)	(5)	(3)
	5×5	95,1	93,7	94,0	94,2	93,6	93,5	93,4
		(3)	(24)	(5)	(3)	(24)	(5)	(3)
	7×7	94,9	93,5	93,8	94,0	93,3	92,9	93,3
		(3)	(26)	(6)	(3)	(25)	(6)	(3)

pay in terms of labeling accuracy (Fig. 3)! This validates the claim of [2] that their geometric approach yields robust flows towards high-quality labelings, despite being overall non-convex.

6   Conclusion

We generalized the transformation of the uncoupled replicator equation from [3] to gradient flows of arbitrary objective functions on any Riemannian manifold. This transformation was then applied to the assignment manifold, to reformulate the gradient flow of [2] for image labeling on a vector space, amenable to numerical integration. This enables the generation of parallel multiplicative update schemes using etablished methods of numerical integration.

A comparison of various update schemes reveales a remarkable efficiency of the adaptive scheme used in [2] regarding *both* required number of iterations and labeling accuracy.

Our further work will study a major extension of the present framework, in order to address the open point: How to cope with the flow *on* the ε-simplex boundary in a *continuous and efficient* manner, similar to the 'interior' flow studied in this paper.

References

1. Absil, P.-A., Mathony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton, Woodstock (2008)
2. Åström, F., Petra, S., Schmitzer, B., Schnörr, C.: Image labeling by assignment. J. Math. Imaging Vis. **58**(2), 211–238 (2017)
3. Ay, N., Erb, I.: On a notion of linear replicator equations. J. Dyn. Differ. Equ. **17**(2), 427–451 (2005)
4. Bergmann, R., Fitschen, J.H., Persch, J., Steidl, G.: Iterative multiplicative filters for data labeling. Int. J. Comput. Vis. 1–19 (2017). http://dx.doi.org/10.1007/s11263-017-0995-9
5. Burbea, J., Rao, C.R.: Entropy differential metric, distance and divergence measures in probability spaces: a unified approach. J. Multivar. Anal. **12**, 575–596 (1982)
6. Hairer, E., Nørsett, S.P., Wanner, G.: Solving Ordinary Differential Equations I. Springer Series in Computational Mathematics, vol. 8, 2nd edn. Springer, Berlin (1993)
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 **155**(2), 155–184 (2015)
8. Lee, J.M.: Introduction to Smooth Manifolds. Springer, New York (2003)