A Fanning Scheme for the Parallel Transport Along Geodesics on Riemannian Manifolds

Maxime Louis, Benjamin Charlier, Paul Jusselin, Susovan Pal, Stanley Durrleman

A FANNING SCHEME FOR THE PARALLEL TRANSPORT ALONG GEODESICS ON RIEMANNIAN MANIFOLDS

MAXIME LOUIS [1,2], BENJAMIN CHARLIER[3,1,2], PAUL JUSSELIN[4,1,2], SUSOVAN PAL[1,2], STANLEY DURRLEMAN[1,2]
[1]INRIA PARIS, ARAMIS PROJECT-TEAM, 75013, PARIS, FRANCE
[2] SORBONNE UNIVERSITÉS, UPMC UNIV PARIS 06, INSERM, CNRS, INSTITUT DU CERVEAU ET DE LA MOELLE (ICM) - HÔPITAL PITIÉ-SALPÊTRIÈRE, BOULEVARD DE L'HÔPITAL, F-75013, PARIS, FRANCE
[3]INSTITUT MONTPELLIÉRAIN ALEXANDER GROTHENDIECK, CNRS, UNIV. MONTPELLIER
[4]CMLA, ENS CACHAN

Abstract. Parallel transport on Riemannian manifolds allows one to connect tangent spaces at different points in an isometric way and is therefore of importance in many contexts, such as statistics on manifolds. The existing methods to compute parallel transport require either the computation of Riemannian logarithms, such as the Schild's ladder, or the Christoffel symbols. The Logarithm is rarely given in closed form, and therefore costly to compute whereas the number of Christoffel symbols explodes with the dimension of the manifold, making both these methods intractable. From an identity between parallel transport and Jacobi fields, we propose a numerical scheme to approximate the parallel transport along a geodesic. We find and prove an optimal convergence rate for the scheme, which is equivalent to Schild's ladder's. We investigate potential variations of the scheme and give experimental results on the Euclidean two-sphere and on the manifold of symmetric positive definite matrices.

Key words. Parallel Transport, Riemannian manifold, Numerical scheme, Jacobi field

1. Introduction.

Riemannian geometry has been long contained within the field of pure mathematics and theoretical physics. Nevertheless, there is an emerging trend to use the tools of the Riemannian geometry in statistical learning to define models for structured data. Such data may be defined by invariance properties, and therefore seen as points in quotient spaces as for shapes, orthogonal frames, or linear subspaces. They may be defined also by smooth inequalities, and therefore as points in open subsets of linear spaces, as for symmetric definite positive matrices, diffeomorphisms or bounded measurements. Such data may be considered therefore as points in a Riemannian manifolds, and analysed by specific statistical approaches [12, 2, 8, 3]. At the core of these approaches lies parallel transport, an isometry which allows the comparison of probability density functions, coordinates or vectors that are defined in the tangent space at different points on the manifold. The inference of such statistical models in practical situations requires therefore efficient numerical schemes to compute parallel transport on manifolds.

The parallel transport of a given tangent vector is defined as the solution of an ordinary differential equation ([6] page 52). In small dimension, this equation is solved using standard numerical schemes. However, this equation requires the computation of the Christoffel symbols whose number explodes with the dimension of the manifold in a combinatorial manner, which makes this approach intractable in realistic situations in statistics.

An alternative is to use the Schild's ladder [1], or its faster version in the case of geodesics the Pole's ladder [5]. These schemes essentially requires the computation of Riemannian exponentials (Exp) and logarithms (Log) at each step. Usually, the computation of the exponential may be done by integrating Hamiltonian equations, and do not raise specific difficulties. By contrast, the computation of the logarithm must often be done by solving an inverse problem ($Exp \circ Log(x) = x$) with the use of an optimization scheme such as a gradient descent. Such optimization schemes are approximate and sensitive so the initial conditions and to hyper-parameters, which leads to additional numerical errors at each step of the scheme. The effects of those numerical errors on the global convergence of the scheme still remain to be studied. When closed formulas exist for the Riemannian logarithm, or in the case of Lie groups, where the Logarithm can be approximated efficiently using the Baker-Campbell-Haussdorff formula (see [4]), the Schild's ladder is an efficient alternative. When this is not the case, it becomes hardly tractable.

Another alternative is to use an equation showing that parallel transport along geodesics may be locally approximated by a well-chosen Jacobi field, up to the second order error. This idea has been suggested in [10] with further credits to [9], but without either a formal definition nor a proof of its convergence. It relies solely on the computations of Riemannian exponentials.

In this paper, we propose a numerical scheme built on this idea, which tries to limit as much as possible the number of operations required to reach a given acuracy. We will prove that this scheme converges at linear speed with the time-step, and that this speed may not be improved without further assumptions on the manifold. Furthermore, we propose an implementation which allows the simultaneous computation of the geodesic and of the transport along this geodesic. Numerical experiments on the 2-sphere and on the manifold of 3-by-3 symmetric positive definite matrices will confirm that the convergence of the scheme is of the same order as the Schild's ladder in practice. Thus, they will show that this scheme offers a compelling alternative to compute parallel transport in high-dimensional manifolds with a control over the numerical errors and the computational cost.

2. Rationale.

2.1. Notations and assumptions. In this paper, we assume that γ is a geodesic defined for all time $t \in [0,1]$ on a manifold $\mathcal{M}$ of finite dimension $n \in \mathbb{N}$ provided with the Riemannian metric g. We denote the Riemannian exponential Exp and ∇ the covariant derivative. For $p \in \mathcal{M}$, $T_p\mathcal{M}$ denotes the tangent space of $\mathcal{M}$ at p. For a vector $w \in T_{\gamma(s)}\mathcal{M}$, for $s, t \in [0,1]$, we denote $\mathrm{P}_{s,t}(w) \in T_{\gamma(t)}\mathcal{M}$ the parallel transport of w from $\gamma(s)$ to $\gamma(t)$. It is the unique solution at time t of the differential equation $\nabla_{\dot\gamma(u)}\mathrm{P}_{s,u}(w) = 0$ for $\mathrm{P}_{s,s}(w) = w$. We also note $\mathrm{J}^w_{\gamma(t)}(h)$ the Jacobi Field emerging from $\gamma(t)$ in the direction $w \in T_{\gamma(t)}\mathcal{M}$, that is:

$$\mathrm{J}^w_{\gamma(t)}(h) = \left.\frac{\partial}{\partial\varepsilon}\right|_{\varepsilon=0} \mathrm{Exp}_{\gamma(t)}(h(\dot\gamma(t) + \varepsilon w)) \in T_{\gamma(t+h)}\mathcal{M}$$

for $h \in \mathbb{R}$ small enough. It verifies the Jacobi equation (see for instance [6] page 111-119):

$$\nabla^2_{\dot\gamma} J^w_{\gamma(t)}(h) + R(J^w_{\gamma(t)}(h), \dot\gamma(h))\dot\gamma(h) = 0 \tag{1}$$

where R is the curvature tensor. We denote $\|\cdot\|_g$ the Riemannian norm on the tangent spaces defined from the metric g, taken at the appropriate point. We use Einstein notations. Throughout the paper, we suppose that there exists a global coordinate system on $\mathcal{M}$ and we note $\Phi : \mathcal{M} \longrightarrow U$ the corresponding diffeomorphism, where U is a subset of $\mathbb{R}^n$. This system of coordinates allows us to define a basis of the tangent space of $\mathcal{M}$ at any point, we note $\left.\frac{\partial}{\partial x^i}\right|_p$ the i-th element of the corresponding basis of $T_p\mathcal{M}$ for any $p \in \mathcal{M}$.

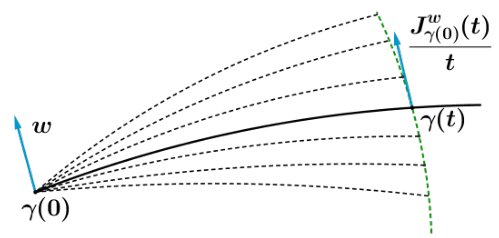

FIGURE 1. *The solid line is the geodesic. The green dotted line is formed by the perturbed geodesics at time t. The blue arrows are the initial vector and its approximated parallel transport at time t.*

We assume that there exists a compact subset K of $\mathcal{M}$ such that $\gamma([0,1]) \subset K$. We also assume that there exists $\eta > 0$ such that injectivity radius of the manifold $\mathcal{M}$ is strictly larger than η.

2.2. The key identity. The numerical scheme that we propose arises from the following identity, which is mentioned in [10]. Figure 1 illustrates the principle.

PROPOSITION 2.1. *For all $t > 0$, and $w \in T_{\gamma(0)}\mathcal{M}$ we have*

$$\mathrm{P}_{0,t}(w) = \frac{\mathrm{J}^w_{\gamma(0)}(t)}{t} + \mathrm{O}\big(t^2\big) \tag{2}$$

Proof. Let $X(t) = P_{0,t}(w)$ be the vector field following the parallel transport equation: $\dot{X}^i + \Gamma^i_{kl} X^l \dot{\gamma}^k = 0$ with $X(0) = w$. In normal coordinates centered at $\gamma(0)$, the Christoffel symbols vanish at $\gamma(0)$ and the equation gives: $\dot{X}^i(0) = 0$. A Taylor expansion of $X(t)$ near $t = 0$ in this local chart then writes:

$$X^i(t) = w^i + \mathrm{O}\big(t^2\big). \tag{3}$$

By definition, the i-*th* normal coordinate of $\mathrm{Exp}_{\gamma(0)}\,(t(v_0 + \varepsilon w))$ is $t(v_0^i + \varepsilon w^i)$. Therefore, the i-*th* coordinate of $\mathrm{J}^w_{\gamma(0)}(t) = \frac{\partial}{\partial \varepsilon}|_{\varepsilon=0} \mathrm{Exp}_{\gamma(0)}(t(\dot{\gamma}(0) + \varepsilon w))$ is tw^i. Plugging this into (3) yields the desired result. ∎

This control on the approximation of the transport by the Jacobi field suggests to divide $[0,1]$ into N intervals $[\frac{k}{N}, \frac{k+1}{N}]$ of length $h = \frac{1}{N}$ for $k = 0, \dots, N-1$ and to approximate the parallel transport of a vector $w \in T_{\gamma(0)}$ from $\gamma(0)$ to $\gamma(1)$ by a sequence of vectors $w_k \in T_{\gamma(\frac{k}{N})}\mathcal{M}$ defined as:

$$\begin{cases} w_0 = w \\ w_{k+1} = N \mathrm{J}^{w_k}_{\gamma(\frac{k}{N})}\left(\dfrac{1}{N}\right) \end{cases} \tag{4}$$

With the control given in the Proposition 2.1, we can expect to get an error of order $\mathrm{O}\big(\frac{1}{N^2}\big)$ at each step and hence a speed of convergence in $\mathrm{O}\big(\frac{1}{N}\big)$ overall. There are manifolds for which the approximation of the parallel transport by Jacobi field is exact e.g. Euclidean space, but in the general case, one cannot expect to get a better convergence rate. Indeed, we show in the next Section that this scheme for the sphere $\mathbb{S}^2$ has a speed of convergence exactly proportional to $\frac{1}{N}$.

2.3. Convergence rate on $\mathbb{S}^2$. In this Section, we assume that one knows the geodesic path $\gamma(t)$ and how to compute any Jacobi fields without numerical errors, and show that the approximation due to Equation (2) alone raises a numerical error at least of order $\mathrm{O}\big(\frac{1}{N}\big)$.

Let $p \in \mathbb{S}^2$ and $v \in T_p\mathbb{S}^2$. (p and v are seen as vectors in $\mathbb{R}^3$). The geodesics are the great circles, which may be written as:

$$\gamma(t) = \mathrm{Exp}_p(tv) = \cos(t|v|)p + \sin(t|v|)\frac{v}{|v|},$$

where $|\cdot|$ is the euclidean norm on $\mathbb{R}^3$. It is straightforward to see that the parallel transport of $w = p \times v$ along $\gamma(t)$ has constant (θ, ϕ) coordinates.

We assume now that $|v| = 1$. Since $w = p \times v$ is orthogonal to v, we have $\frac{\partial}{\partial \varepsilon}\big|_{\varepsilon=0} |v + \varepsilon w| = 0$. Therefore:

$$\begin{aligned} \mathrm{J}_p^w(t) &= \frac{\partial}{\partial \varepsilon}|_{\varepsilon=0}\mathrm{Exp}_p(t(v+\varepsilon w)) \\ &= \frac{\partial}{\partial \varepsilon}|_{\varepsilon=0}\left(\cos(t|v+\varepsilon w|)p + \sin(t|v+\varepsilon w|)\frac{v+\varepsilon w}{|v+\varepsilon w|}\right) \\ &= \sin(t)w \end{aligned}$$

which does not depend on p. We have $\mathrm{J}^w_{\gamma(t)}(t) = \sin(t)w$. Consequently, the sequence of vectors w_k built by the iterative process described in Equation (4) verifies $w_{k+1} = N w_k \sin\left(\frac{1}{N}\right)$ for $k = 0, ..., N-1$, and $w_N = w_0 N \sin\left(\frac{1}{N}\right)^N$. In tangent space coordinates, $P_{0,1}(w_0) = w_0$, so that the numerical error, measured in those tangent space coordinates, is proportional to $w_0\left(1 - \left(\frac{\sin(1/N)}{1/N}\right)^N\right)$. We have:

$$\left(\frac{\sin(1/N)}{1/N}\right)^N = \exp\left(N \log\left(1 - \frac{1}{6N^2} + \mathrm{o}\big(1/N^2\big)\right)\right) = 1 - \frac{1}{6N} + \mathrm{o}\big(\frac{1}{N}\big)$$

yielding:

$$\frac{|w_N - w_0|}{|w_0|} \propto \frac{1}{6N} + \mathrm{o}\big(\frac{1}{N}\big).$$

It shows a case where the bound $\frac{1}{N}$ is reached.

3. The numerical scheme.

3.1. The algorithm.

Unless the metric has some nice properties, there are no closed forms expressions for the geodesics and the Jacobi fields. Hence, in most practical cases, these quantities also need to be computed using numerical methods.

Computing geodesics. In order to avoid the computation of the Christoffel symbols, we propose to integrate the first-order Hamiltonian equations to compute geodesics (see [11]). Let $x(t) = (x_1(t), \ldots, x_d(t))^T$ be the coordinates of $\gamma(t)$ in a given local chart, and $\alpha(t) = (\alpha_1(t), \ldots, \alpha_d(t))^T$ be the coordinates of the momentum $g(\gamma(t))\dot{\gamma}(t) \in T^*_{\gamma(t)}\mathcal{M}$ in the same local chart. We have then:

$$\begin{cases} \dot{x}(t) = K(x(t))\alpha(t) \\ \dot{\alpha}(t) = -\frac{1}{2}\nabla_x\left(\alpha(t)^T K(x(t))\alpha(t)\right) \end{cases}, \tag{5}$$

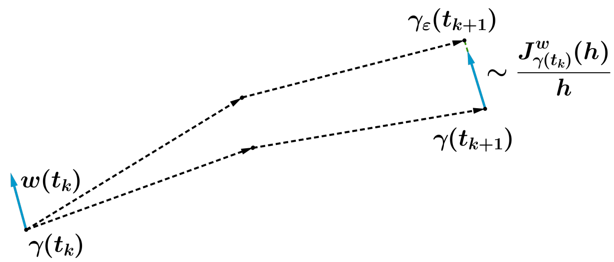

FIGURE 2. *One step of the numerical scheme. The dotted arrows represent the steps of the Runge-Kutta integrations for the main geodesic γ and for the perturbed geodesic γ^ε. The blue arrows are the initial w and the obtained approximated transport using equation (6).*

where $K(x(t))$, a d-by-d matrix, is the inverse of the metric g expressed in the local chart. We will see that to ensure the convergence of the scheme we must use a Runge-Kutta scheme of order at least 2 to integrate this equation, for which the error is in $\mathrm{O}\big(\frac{1}{N^2}\big)$.

Computing $\mathrm{J}^w_{\gamma(t)}(h)$. The Jacobi field may be approximated with a numerical differentiation from the computation of a perturbed geodesic γ^ε with initial position $\gamma(t)$ and initial velocity $\dot\gamma + \varepsilon w$ where ε is a small parameter:

$$\mathrm{J}^w_{\gamma(t)}(h) \simeq \frac{\mathrm{Exp}_{\gamma(t)}(h(\dot\gamma(t) + \varepsilon w)) - \mathrm{Exp}_{\gamma(t)}(h(\dot\gamma(t))}{\varepsilon}, \tag{6}$$

where the Riemannian exponential may be computed by integration of the Hamiltonian equations (5) over the time interval $[t, t+h]$ starting at point $\gamma(t)$, see Figure 2. We will also see that, in general, a choice for ε ensuring a $\mathrm{O}\big(\frac{1}{N}\big)$ order of convergence is $\varepsilon = \frac{1}{N}$.

The algorithm. Let $N \in \mathbb{N}$. We divide $[0,1]$ into N intervals $[t_k, t_{k+1}]$, and initialize with $\gamma_0 = \gamma(0)$, $\dot\gamma_0 = \dot\gamma(0)$ and $w_0 = w$. The algorithm we propose consists in iteratively computing, at step k :

(i) The momentum in the cotangent space corresponding to the vector w_k: $\beta_k = K(\gamma_k)w_k$

(ii) The new point on the main geodesic γ_{k+1}, by integration of the Hamiltonian equations using a second-order Runge-Kutta method.

(iii) The perturbed geodesic starting at γ_k with initial tangent vectors $\dot\gamma_k + \varepsilon w_k$ at time h, that we denote γ^ε_{k+1} using a second-order Runge-Kutta method.

(iv) The estimated parallel transport before renormalization :

$$\hat w_{k+1} = \frac{\gamma^\varepsilon_{k+1} - \gamma_{k+1}}{\varepsilon}$$

(v) The new estimated parallel transport :

$$w_{k+1} = \alpha_k \hat w_{k+1} + \beta_k \dot\gamma_{k+1}$$

where α_k and β_k are normalization factors ensuring $\|w(t_{k+1})\|_g = \|w(t_0)\|_g$ and $g(w_{k+1}, \dot\gamma_{k+1}) = g(w_0, \dot\gamma_0)$: those quantitites should be conserved during the transport. This comes at a small cost, and we will see in Proposition 4.2 that it allows to put a uniform bound on the approximation of the transport by the Jacobi field.

Figure 2 illustrates the principle. A complete pseudo-code is given in appendix A. It is remarkable that we can substitute the computation of the Jacobi Field with only four calls to the hamiltonian equations (5) at each step, including the calls necessary to compute the main geodesic. Note however that the (i) step of the algorithm requires to solve a linear system, which is an operation whose cost increases with the dimension, in a polynomial manner.

3.2. Order of the approximations and quantity conservations. As we will see below, the orders of the different approximations presented above are optimal in the sense that they are minimal to ensure linear convergence of the scheme. We could increase the order of the Runge-Kutta integration in the steps (ii) or (iii), or increase the order of the finite difference approximation of the derivative in step (iii) e.g. by computing two perturbed geodesics and using a central finite difference:

$$\mathrm{J}^w_{\gamma(t)}(h) \simeq \frac{\mathrm{Exp}(h(\dot\gamma(t)+\varepsilon w)) - \mathrm{Exp}(h(\dot\gamma(t)-\varepsilon w))}{2\varepsilon},$$

which is of order 2 instead of the assymetric first-order approximation proposed here. This method requires 6 calls to the Hamiltonian equations, instead of 4. We will study both of these in Section 6 to identify the most cost-effective method to reach a given precision.

Remark. To ensure the conservations of both these quantities, we can either solve the linear system to find α and β at step (v), or we can alternatively split w into two components : $w_\parallel = \frac{g(v,w)}{\|v\|_g} v$ being the component of w parallel to the initial velocity and $w_\perp$ the orthogonal component, transport them separately while ensuring simple renormalizations and combining the results in the end. It is an alternative with a different implementation that might be convenient in some cases.

3.3. The convergence Theorem. We obtained the following convergence result, guaranteeing a linear decrease of the error with the size of the step h.

THEOREM 3.1. *Let $N \in \mathbb{N}$. Let $w \in T_{\gamma(0)}\mathcal{M}$. We denote $\delta_k = \|P_{0,t_k}(w) - \tilde w_k\|_2$ where $\tilde w_k$ is the approximate value of the parallel transport of w along γ at time t_k and where the 2-norm is taken in the coordinates of our global chart. We note ε the parameter used in the step (iii) and $h = \frac{1}{N}$ the size of the step used of the Runge-Kutta approximate solution of the geodesic equation.*
With the hypotheses stated in Section 2.1, if we take $\varepsilon = \frac{1}{N}$, then we have:

$$\delta_N = O(\frac{1}{N}).$$

We will see in the proof and in the numerical experiments that choosing $\varepsilon = h$ is a recommended choice for the size of the step in the differentiation of the perturbed geodesics, that further decreasing ε has no visible effect on the accuracy of the estimation and that choosing a larger ε lowers the quality of the approximation.

Note that our result controls the 2-norm of the error in the global system of coordinates, but not directly the metric norm in the tangent space at $\gamma(1)$. This is due to the fact that our knowledge of the main geodesic is approximate, with a residual error preventing us from using the metric g at $\gamma(1)$ as a measure of the error. However, studying the convergence in the global system of coordinates corresponds to a relevant notion of convergence, since the error on the approximation of $\gamma(1)$ is of order $\mathrm{O}(h^2)$ and the metric is smooth.

Before giving a proof of this theorem in Section 5, we prove some lemmas allowing uniform controls on the different sources of error in the numerical scheme. In Section 4.1, we prove an intermediate results allowing uniform controls on norms of tensors, in Section 4.2, we prove a stronger result than Proposition 2.1, with stronger hypotheses and in Section 4.3, we prove a result allowing to control the accumulation of the error.

4. Proofs of the lemmas.

4.1. A lemma to change coordinates. We recall that we suppose the geodesic contained within a compact subset K of the manifold. We start with a result controlling the norms of change-of-coordinates matrices. Let p in $\mathcal{M}$ and $q \in \mathcal{M}$ within the radius of the exponential map at p. We consider two basis on $T_p\mathcal{M}$: one defined from the global system of coordinates, that we note , and another made of the normal coordinates (defined from the global system of coordinates Φ) centered at p, that we note B_p^Ψ. We can therefore define $\Lambda(p,q)$ as the change-of-coordinates matrix between B_p^Φ and B_p^Ψ. The operators norms $|||\cdot|||$ of these matrices are bounded over K in the following sense :

LEMMA 4.1. *There exists $L \geq 0$ such that for all $p \in K$, for all $q \in K$ such that $q = \mathrm{Exp}_p(v)$ for some $v \in T_p\mathcal{M}$, with $\|v\|_g \leq \frac{\eta}{2}$ then :*

$$|||\Lambda(p,q)||| \leq L$$

and

$$|||\Lambda^{-1}(p,q)||| \leq L.$$

Proof. Let $p \in \mathcal{M}$. We identify $T_p\mathcal{M}$ with $\mathbb{R}^n$ to get a norm $\|\cdot\|_{g(p)}$ on $\mathbb{R}^n$. This norm is equivalent to the 2-norm $\|\cdot\|_2$ so that there exists $A > 0$ such that for all $v \in \mathbb{R}^n$, $\|v\|_2 \leq A\|v\|_{g(p)}$. Because K is compact and g varies smoothly, there exists a constant $A' > 0$ which makes this majoration valid at any point, i.e. such that for all $p \in \mathcal{M}$, for all $v \in \mathbb{R}^n$, we have :

$$\|v\|_2 \leq A'\|v\|_g \tag{7}$$

We note $B(0, \frac{\eta}{2A'})$ the closed ball of radius $\frac{\eta}{2A'}$ in $(\mathbb{R}^n, \|\cdot\|_2)$. Let $(p,v) \in K \times B(0, \frac{\eta}{A'})$. We note $q = \mathrm{Exp}_p(v)$. The application $\Lambda : (p,v) \to |||\Lambda(p,v)|||$ is smooth, because the change of basis matrices smoothly depend on the metric g and on the positions of p and q. Moreover, Λ is defined on a compact set and hence reaches its maximum $L \geq 0$. Thanks to the upper bound in (7), when v spans $B(0, \frac{\eta}{2A'})$ in $(\mathbb{R}^n, \|\cdot\|_2)$, it does stay within $B(0, \frac{\eta}{2})$ in $(T_p\mathcal{M}, \|\cdot\|_g)$ so that the bound L on Λ is valid for all $p \in \mathcal{M}$ and for all q such that $q = \mathrm{Exp}_p(v)$ with $\|v\|_g \leq \frac{\eta}{2}$. We proceed similarly for Λ^{-1}. ∎

This lemma allows us to translate any bound on the components of a tensor in the global system of coordinates into a bound on the components of the same tensor in any of the normal systems of coordinates centered at a point of the geodesic, and *vice versa*.

4.2. A stronger version of Proposition 2.1. From there, we can prove a stronger version of Proposition 2.1. We use here the assumption that the manifold has a strictly positive injectivity radius η on K.

Proposition 4.2. *There exists $A \geq 0$ such that for all $t \in [0,1[$, for all $w \in T_{\gamma(t)}\mathcal{M}$ and for all $h < \max\left(\frac{\eta}{\|\dot\gamma(t)\|_g}, 1-t\right)$:*

$$\left\| P_{t,t+h}(w) - \frac{\mathrm{J}^w_{\gamma(t)}(h)}{h} \right\|_g \leq Ah^2 \|w\|_g.$$

Proof. Let $t \in [0,1[$, $w \in T_{\gamma(t)}\mathcal{M}$ and $h < \max\left(\frac{\eta}{\|\dot\gamma(t)\|_g}, 1-t\right)$ i.e. such that $J^w_{\gamma(t)}(h)$ is well defined. The following identity, satisfied for any smooth vector field V on $\mathcal{M}$:

$$\nabla^k_{\dot\gamma} V(\gamma(t)) = \frac{\mathrm{d}^k}{\mathrm{d}h^k} P^{-1}_{t,t+h}(V(\gamma(t+h)) \tag{8}$$

which will be proved in Appendix B.1 provides us with a way to compute the successive derivatives of $P^{-1}_{t,t+h}(J^w_{\gamma(t)}(h))$.

We have $J^w_{\gamma(t)}(0) = 0$, $\nabla_{\dot\gamma} J^w_{\gamma(t)}(0) = w$, $\nabla^2_{\dot\gamma} J^w_{\gamma(t)}(0) = -R(J^w_{\gamma(t)}(0), \dot\gamma(0))\dot\gamma(0) = 0$ using equation (1) and finally:

$$\begin{aligned} \|\nabla^3_{\dot\gamma} J^w_{\gamma(t)}(h)\|_g &= \|\nabla_{\dot\gamma}(R)(J^w_{\gamma(t)}(h), \dot\gamma(h))\dot\gamma(h) + R(\nabla_{\dot\gamma} J^w_{\gamma(t)}(h), \dot\gamma(h))\dot\gamma(h)\|_g \\ &\leq \|\nabla_{\dot\gamma} R\|_\infty \|\dot\gamma(h)\|_g^2 \|J^w_{\gamma(t)}(h)\|_g + \|R\|_\infty \|\dot\gamma(h)\|^2_g \|\nabla_{\dot\gamma} J^w_{\gamma(t)}(h)\|_g, \end{aligned} \tag{9}$$

where the ∞-norms, taken over the geodesic and the compact K, are finite because the curvature and its derivatives are bounded. In normal coordinates centered at $\gamma(t)$, we have $J^w_{\gamma(s)}(h)^i = hw^i$. Therefore, if we note $g_{ij}(\gamma(t+h))$ the components of the metric in the normal coordinates, we get:

$$\|J^w_{\gamma(t)}(h)\|^2_g = h^2 g_{ij}(\gamma(t+h)) w^i w^j.$$

To obtain an upper bound for this term which does not depend on t, we note that the coefficients of the metric in the global coordinate system are bounded on K. Using the Lemma 4.1, we get a bound into a bound $M \geq 0$ valid on all the normal system of coordinates centered at a point of the geodesic, so that:

$$\|J^w_{\gamma(t)}(g)\|_g \leq hM\|w\|_2.$$

By equivalence of the norms as seen in the lemma (4.1), and because g varies smoothly, there exists $N \geq 0$ such that:

$$\|J^w_{\gamma(t)}(g)\|_g \leq hMN\|w\|_g \tag{10}$$

where the dependence of the majoration on t has vanished, and the result stays valid for all $h < \max\left(\frac{\eta}{\|\dot\gamma(t)\|_g}, 1-t\right)$ and all w. Similarly, there exists $C > 0$ such that :

$$\|\nabla_{\dot\gamma} J^w_{\gamma(s)}(h)\| \leq C\|w\|_g, \tag{11}$$

at any point and for any $h < \max\left(\frac{\eta}{\|\dot\gamma(t)\|_g}, 1-t\right)$. Gathering equations (9), (10), (11) , we get that there exists a constant $A \geq 0$ which does not depend on t, h or w such that:

$$\left\|\nabla^3_{\dot\gamma} J^w_{\gamma(s)}(h)\right\|_g \leq A\|w\|_g. \tag{12}$$

Now using equation (8) with $V = J^w_{\gamma(t)}$ and a Taylor's formula, we get :

$$J^w_{\gamma(t)}(h) = hP_{t,t+h}(w) + P_{t,t+h}(r(w,h))$$

where we noted r the remainder of the expansion. Therefore :

$$\left\| \frac{J^w_{\gamma(t)}(h)}{h} - P_{t,t+h}(w) \right\|_g = \|P_{t,t+h}(r(w,h))\|_g.$$

Now, because the parallel transport is an isometry and thanks to the equation (12):

$$\left\| \frac{J^w_{\gamma(t)}(h)}{h} - P_{t,t+h}(w) \right\|_g \le \frac{A}{6} h^2 \|w\|_g.$$

□

4.3. A lemma to control the accumulation of the error. At every step of the scheme, we compute a Jacobi field from an approximate value of the transported vector. We need to control the error made with this computation from an already approximate vector. We provide a control on the 2-norm of the corresponding error, in the global system of coordinates.

LEMMA 4.3. *There exists $B \ge 0$ such that for all $t \in [0,1[$, for all $w_1, w_2 \in T_{\gamma(t)}\mathcal{M}$, for all $h \le \frac{\eta}{\|\dot\gamma(t)\|_g}$ small enough, we have :*

$$\left\| \frac{J^{w_1}_{\gamma(t)}(h) - J^{w_2}_{\gamma(t)}(h)}{h} \right\|_2 \le (1 + Bh)\|w_1 - w_2\|_2. \quad (13)$$

Proof. Let $t \in [0,1[$ and $h \in [0, 1-t]$. We note $p = \gamma(t)$, $q = \gamma(t+h)$. We use the exponential map to get normal coordinates on a neighborhood of V of p from the basis $\frac{\partial}{\partial x^i}\big|_p$ of $T_p\mathcal{M}$. Let's note $\frac{\partial}{\partial y^i}\Big|_r$ the corresponding basis on the tangent space at any point r of V. Let $w_1, w_2 \in T_p\mathcal{M}$ and note w_i^j for $i \in \{1,2\}$, $j \in \{1, ..., n\}$ the coordinates in the global system. By definition, the basis $\left(\frac{\partial}{\partial y^k}\Big|_p\right)$ and the basis $\left(\frac{\partial}{\partial x^k}\big|_p\right)$ coincide, and in particular, for $i \in \{1,2\}$:

$$w_i = (w_i)^k \left.\frac{\partial}{\partial x^k}\right|_p = (w_i)^k \left.\frac{\partial}{\partial y^k}\right|_p.$$

If $i \in \{1,2\}$, $j \in \{1, ..., n\}$, the j-th coordinate of $J^{w_i}_{\gamma(t)}(h)$ in the basis $\left(\frac{\partial}{\partial y^i}\Big|_q\right)_{i=1,\dots,n}$ is:

$$J^{w_i}_{\gamma(t)}(h)^j = \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} (\exp_p(h(v + \varepsilon w_i)))^j = \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} (h(v + \varepsilon w_i))^j = hw^j.$$

Let $\Lambda(\gamma(t+h), \gamma(t))$ be the change-of-coordinate matrix of $T_{\gamma(t+h)}$ from the basis $\left(\frac{\partial}{\partial y^k}\Big|_q\right)$ to the basis $\left(\frac{\partial}{\partial x^k}\big|_q\right)$. Λ varies smoothly with t and h, and is the identity when $h = 0$. Hence, we can write an expansion :

$$\Lambda(\gamma(t+h), \gamma(t)) = Id + hV(t) + O(h^2)$$

The second order term depends on the second derivative of Λ with respect to h. Restricting ourselves to a compact subset, as in the Lemma 4.1, we get a uniform bound on the norm of this second derivative thus getting a control on the operator norm of $\Lambda(\gamma(t+h), \gamma(t))$, that we can write, for h small enough :

$$|||\Lambda(\gamma(t+h), \gamma(t))||| \leq (1 + Bh)$$

where B is a positive constant which does not depend on h or t. Now we get :

$$\left\| \frac{J^{w_1}_{\gamma(t)}(h) - J^{w_2}_{\gamma(t)}(h)}{h} \right\|_2 = \|\Lambda(\gamma(t+h), \gamma(t))(w_1 - w_2)\|_2 \leq (1 + Bh)\, \|w_1 - w_2\|_2$$

which is the desired result. ∎

5. Proof of the convergence Theorem 3.1.

Proof. Let $k \in \mathbb{N}$. We build an upper bound on the error δ_{k+1} from δ_k. We have :

$$\begin{aligned}
\delta_{k+1} &= \|w_{k+1} - \tilde{w}_{k+1}\|_2 \\
&\leq \underbrace{\left\| w_{k+1} - \frac{\mathrm{J}^{w_k}_{\gamma_k}(h)}{h} \right\|_2}_{(1)} + \underbrace{\left\| \frac{\mathrm{J}^{w_k}_{\gamma_k}(h)}{h} - \frac{\mathrm{J}^{\tilde{w}_k}_{\gamma_k}(h)}{h} \right\|_2}_{(2)} \\
&\quad + \underbrace{\left\| \frac{\mathrm{J}^{\tilde{w}_k}_{\gamma_k}(h)}{h} - \frac{\mathrm{J}^{\tilde{w}_k}_{\tilde{\gamma}_k}(h)}{h} \right\|_2}_{(3)} + \underbrace{\left\| \frac{\mathrm{J}^{\tilde{w}_k}_{\tilde{\gamma}_k}(h)}{h} - \frac{\tilde{\mathrm{J}}^{\tilde{w}_k}_{\tilde{\gamma}_k}(h)}{h} \right\|_2}_{(4)}
\end{aligned}$$

where
- $\tilde{\gamma}$ is the approximation of the geodesic coordinates at step k.
- $w_k = P_{0,t_{k+1}}(w)$ is the exact parallel transport.
- $\tilde{w}_k$ is its approximation at step k
- $\tilde{\mathrm{J}}$ is the approximation of the Jacobi field computed with finite difference.
- $\mathrm{J}^{\tilde{w}_k}_{\tilde{\gamma}_k}(h)$ is the Jacobi field computed with the approximations $\tilde{w}$, $\tilde{\gamma}$ and $\tilde{\dot{\gamma}}$.

We control each of these terms.

(1). This is the intrinsic error when using the Jacobi field. We showed in Proposition 4.2 that for h small enough :

$$\left\| P_{t_k,t_{k+1}}(w_k) - \frac{\mathrm{J}^{w_k}_{\gamma(k)}(h)}{h} \right\|_{g(\gamma(t_{k+1}))} \leq\ Ah^2 \|w_k\|_g = Ah^2 \|w_0\|_g$$

Now, since g varies smoothly and by equivalence of the norms, there exists $A' > 0$ such that :

$$\left\| P_{t_k,t_{k+1}}(w_k) - \frac{\mathrm{J}^{w_k}_{\gamma(k)}(h)}{h} \right\|_2 \leq A' h^2 \|w_0\|_g \tag{14}$$

(2). We showed in Section 4.3 below that for h small enough:

$$\left\| \frac{\mathrm{J}^{w_k}_{\gamma(t_k)}(h)}{h} - \frac{\mathrm{J}^{\tilde{w}_k}_{\gamma(t_k)}(h)}{h} \right\|_2 \leq (1 + Bh)\delta_k \tag{15}$$

(3). This term measures the error linked to our approximate knowledge of the geodesic γ. It is proved in Appendix B.2 that there exists a constant $C > 0$ which does not depend on k or h such that :

$$\left\| \frac{\mathrm{J}^{\tilde{w}_k}_{\gamma_k}(h)}{h} - \frac{\tilde{\mathrm{J}}^{\tilde{w}_k}_{\tilde{\gamma}_k}(h)}{h} \right\|_2 \leq Ch^2 \tag{16}$$

(4). This is the difference between the analytical computation of J and its approximation. It is proved in Appendix B.3 and B.4 that if we use a Runge-Kutta method or order 2 to compute the geodesic equations and a second-order method to compute the Jacobi field, or if we use a single perturbed geodesic and a first-order method to compute the Jacobi field, there exists $D \geq 0$ which does not depend on k such that :

$$\left\| \frac{\mathrm{J}^{\tilde{w}_k}_{\gamma(t_k)} - \tilde{\mathrm{J}}^{\tilde{w}_k}_{\gamma(t_k)}}{h} \right\|_2 \leq D(h^2 + \varepsilon h)\|w_0\|_g. \tag{17}$$

Note that D does not depend on k since we renormalize $\tilde{w}$ at each step, thus gaining a control on the norm which is used in Section B.3 and B.4.

Gathering equations (14), (15), (16) and (17), there exists a constant $F > 0$ such that for all k:

$$\delta_{k+1} \leq (1 + Ah)\delta_k + F(h^2 + h\varepsilon).$$

Combining those inequalities for $k = 1, ..., N$, we obtain a geometric series whose sum yields:

$$\delta_N \leq \frac{F(h^2 + h\varepsilon)}{Ah}(1 + Ah)^{N+1}$$

Here we see that choosing $\varepsilon = h$ yields an optimal rate of convergence : choosing a larger value deteriorates the accuracy of the scheme while choosing a lower value still yields a $\frac{1}{N}$ error. Setting $\varepsilon = h$ and recalling that $h = \frac{1}{N}$:

$$\delta_N \leq \frac{2F}{AN}(1 + \frac{A}{N})^{N+1} = \frac{2F}{AN}(\exp(A) + \mathrm{o}(\frac{1}{N})$$

Eventually, there exists $G > 0$ such that, for $N \in \mathbb{N}$ large enough:

$$\delta_N \leq \frac{G}{N}.$$ ∎

It seems that choosing a lower value or ε could improve the performance, however the numerical experiments showed that the accuracy of the differentiation of J seems to be quickly saturated, and the other approximations become limiting.

6. Numerical experiments.

6.1. Setup. We implemented the numerical scheme on simple manifolds where the paralllel transport is known in a closed form, allowing us to evaluate the numerical error [1]. We present two examples :

[1] A modular Python version of the code is available here: https://gitlab.icm-institute.org/maxime.louis/parallel-transport

- $\mathbb{S}^2$: in spherical coordinates (θ, ϕ) the metric is $g = \begin{pmatrix} 1 & 0 \\ 0 & \sin(\theta)^2 \end{pmatrix}$. We gave expressions for geodesics and parallel transport in Section 2.3.
- The set of 3×3 symmetric positive-definite matrices SPD(3). The tangent space at any points of this manifold is the set of symmetric matrices. In [2], the authors endow this space with the affine-invariant metric: for $\Sigma \in \mathrm{SPD}(3)$, $V, W \in \mathrm{Sym}(3)$:

$$g_\Sigma(V, W) = \mathrm{tr}(\Sigma^{-1} V \Sigma^{-1} W)$$

Through an explicit computation of the christoffel symbols, they derive explicit expressions for any geodesic $\Sigma(t)$ starting at $\Sigma_0 \in \mathrm{SPD}(3)$ with initial tangent vector $X \in \mathrm{Sym}(3)$:

$$\Sigma(t) = \Sigma_0^{\frac{1}{2}} \exp(tX) \Sigma_0^{\frac{1}{2}}$$

where $\exp : \mathrm{Sym}(3) \to \mathrm{SPD}(3)$ is the matrix exponentiation. Deriving an expression for the parallel transport can also be done using the explicit Christoffel symbols, see [7]. If $\Sigma_0 \in \mathrm{SPD}(3)$ and $X, W \in \mathrm{Sym}(3)$, then :

$$P_{0,t}(W) = \exp(\frac{t}{2} X \Sigma_0^{-1}) W \exp(\frac{t}{2} \Sigma_0^{-1} X)$$

The code for this numerical scheme can be written in a generic way and used for any manifold by specifying the Hamiltonian equations and the metric.

Remark. Note that even though the computation of the gradient of the inverse of the metric with respect to the position, $\nabla_x K$, is required to integate the Hamiltonian equations (5), $\nabla_x K$ can be computed from the gradient of the metric using the fact that any smooth map $M : \mathbb{R} \to GL_n(\mathbb{R})$ verifies $\frac{\mathrm{d}M^{-1}}{\mathrm{d}t} = -M^{-1} \frac{\mathrm{d}M}{\mathrm{d}t} M^{-1}$. This is how we proceeded for SPD(3): it spares some potential difficulties if one does not have access to analytical expressions for the inverse of the metric.

6.2. Results. Errors measured in the chosen system of coordinates confirm the linear behavior in both cases, as shown on Figures 3 and 4.

We assessed the effect of a higher order for the Runge-Kutta scheme in the integration of geodesics. Using a fourth order method increases the accuracy of the transport in both cases, by a factor 2.3 in the single geodesic case. A fourth order method is twice as expensive as a second order method in terms of number of calls to the Hamiltonian equations, hence in this case it is the most efficient way to reach a given accuracy.

We also investigated the effect of enforcing the conservations of the norm and of the scalar product with the velocity, as discussed in 3.2. Doing so yields an exact transport for the sphere, because it is of dimension 2, and a dramatically improved transport of the same order of convergence for SPD(3) (see Figure 4). The complexity of this operation is very low, and we recommend to always use it. It can be expected however that the effect of the enforcement of these conservations will lower as the dimension increases, since it only fixes two components of the transported vector.

We also confirmed numerically that without a second-order method to integrate the geodesic equations, the scheme does not converge.

Finally, using two geodesic to compute a central-finite difference for the Jacobi Field is 1.5 times more expensive than using a single geodesic, in terms of number of

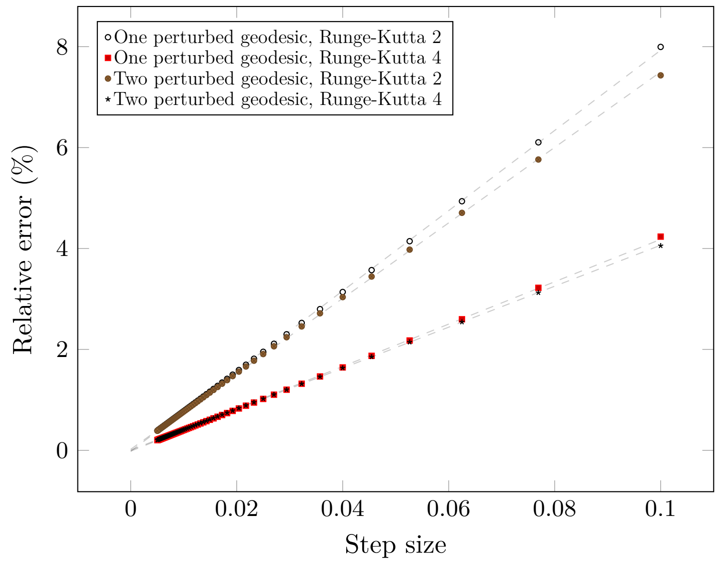

FIGURE 3. *Relative error for the 2-Sphere in defferent settings, as functions of the step size, with initial point, velocity and initial w kept constant. The dotted lines are linear regressions of the measurements.*

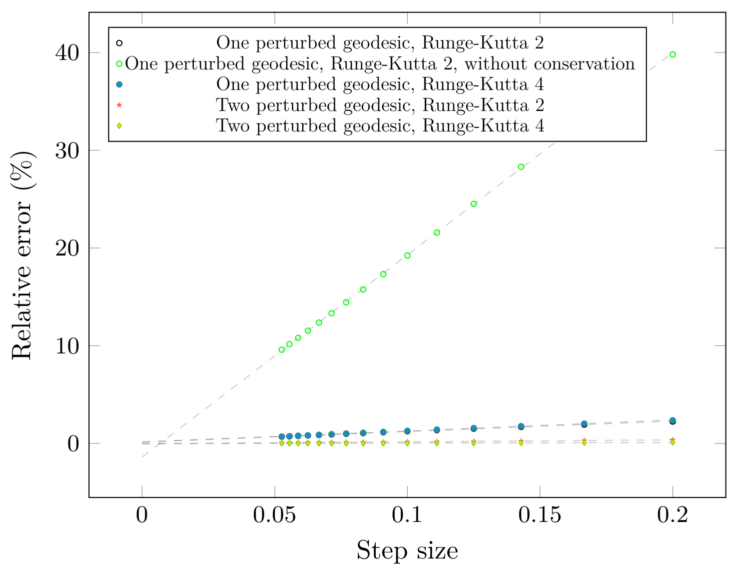

FIGURE 4. *Relative errors for* $SPD(3)$ *in different settings, as functions of the step size, with initial point, velocity and initial w kept constant. The dotted lines are linear regressions.*

calls to the Hamiltonian equations, and it is therefore more efficient to compute two perturbed geodesics in the case of the symmetric positive-definite matrices.

6.3. Comparison with the Schild's ladder. We compared the relative errors of the fanning scheme with the other Christoffel-less method : the Schild's ladder. We implemented the Schild's ladder on the sphere, and compare the relative errors of both schemes on a same geodesic and vector. We chose this vector to be orthogonal to the velocity, since the transport with the Schild's ladder is exact if the transported vector is colinear to the velocity. We use a closed form expression for the Riemannian logarithm in the Schild's ladder, and closed form expressions for the geodesic. The results are given in Figure 5. The fanning scheme is 1.6 times more accurate.

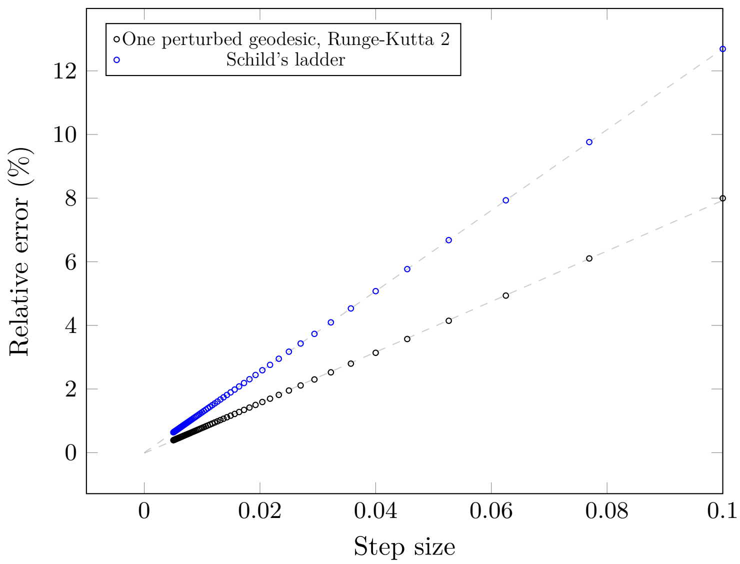

FIGURE 5. *Relative error of the Schild's ladder scheme compared to the fanning scheme (double geodesic, Runge-Kutta 2) proposed here, in the case of* $\mathbb{S}^2$.

The constants in the speed of convergence don't differ much.

7. Conclusion.

We proposed a new method, the fanning scheme, to compute parallel transport along a geodesic on a Riemannian manifold using Jacobi Fields. At variance with the Schild's ladder, this method does not require the computation of Riemannian logarithms, which are in a lot of cases not given in closed form and potentially hard to approximate. We proved that the error of the scheme is of order $O\left(\frac{1}{N}\right)$ where N is the number of discretization steps, and that it cannot be improved in the general case, yielding the same convergence rate as the Schild's ladder. Note also that, to the best of our knowledge, no convergence result is available for the Schild's ladder when extra approximations, which are often necessary, are made –e.g. approximate Riemannian logarithm through gradient descent or using the Baker-Haussdorf-Campbell formula. We also showed that only four calls to the Hamiltonian equations are necessary at each step to provide a satisfying approximation of the transport, two of them being used to compute the main geodesic. We confirmed the rate of convergence numerically, and showed empirically that ensuring the conservations of the norm and of the scalar product with the velocity can yield significative improvements to the approximation, although this fact still needs to be confirmed in high dimensions.

A limitation of this scheme is to only be applicable when parallel transporting along geodesics, and an extension to a more general family of curves would be an interesting perspective. Besides, the Hamiltonian equations are expressed in the cotangent space whereas the velocity lies in the tangent space. Going back and forth from cotangent to tangent space at each iteration can be costly : it typically requires a matrix multiplication, and potentially the inversion of the metric. In very high dimensions this might limit the performances of the scheme.

Acknowledgements. This work has been partially funded by the European Research Council (ERC) under grant agreement No 678304, European Union's Horizon 2020 research and innovation programme under grant agreement No 666992, and the program "Investissements d'avenir" ANR-10-IAIHU-06.

Appendix A. Pseudo-code for the algorithm. We give a pseudo-code description of the numerical scheme. We note G the metric.

function ParallelTransport(x_0, α_0, w_0, N)
▷ x_0 coordinates of $\gamma(0)$
▷ α_0 coordinates of $G(\gamma(0))\dot\gamma(0) \in T^*_{\gamma(0)}\mathcal{M}$
▷ w_0 coordinates of $w \in T_{\gamma(0)}\mathcal{M}$
▷ N number of time-steps
    $h = 1/N$, $\varepsilon = 1/N$
    for $k = 0, \dots, (N-1)$ **do**
▷ integration of the main geodesic
        $x_{k+\frac{1}{2}} = x_k + \frac{h}{2} v_k$
        $\alpha_{k+\frac{1}{2}} = \alpha_k + \frac{h}{2} \mathrm{F}(x_k, \alpha_k)$
        $x_{k+1} = x_k + h\mathrm{V}(x_{k+\frac{1}{2}}, \alpha_{k+\frac{1}{2}})$
        $\alpha_{k+1} = \alpha_k + h\mathrm{F}(x_{k+\frac{1}{2}}, \alpha_{k+\frac{1}{2}})$
▷ perturbed geodesic equation in the direction w_k
        $\beta_k = \mathrm{K}(x_k)^{-1} w_k$
        $\alpha_k^\varepsilon = \alpha_k + \varepsilon \beta_k$
        $x^\varepsilon_{k+\frac{1}{2}} = x_k + \frac{h}{2}(v_k + \varepsilon w_k)$
        $\alpha^\varepsilon_{k+\frac{1}{2}} = \alpha^\varepsilon_k + \frac{h}{2} \mathrm{F}(x_k, \alpha^\varepsilon_k)$
        $x^\varepsilon_{k+1} = x^\varepsilon_k + h\mathrm{V}(x^\varepsilon_{k+\frac{1}{2}}, \alpha^\varepsilon_{k+\frac{1}{2}})$

        $J_{k+1} = \frac{x^\varepsilon - x_{k+1}}{\varepsilon}$ ▷ Jacobi field by finite differences
▷ Conserve quantities
        $v_{k+1} = \mathrm{V}(x_{k+1}, \alpha_{k+1})$
        Solve for a, b :
        $G(w_0, w_0) = G(aJ_{k+1} + bv_{k+1}, aJ_{k+1} + bv_{k+1})$,
        $G(v_0, w_0) = G(aJ_{k+1} + bv_{k+1}, v_{k+1})$
        $w_{k+1} = aJ_{k+1} + bv_{k+1}$ ▷ parallel transport
    end for
    return x_N, α_N, w_N
▷ x_N approximation of $\gamma(1)$
▷ α_N approximation of $G(\gamma(1))\dot\gamma(1)$
▷ w_N approximation of $\mathrm{P}_{\gamma(0),\gamma(1)}(w_0)$
end function

function $\mathrm{V}(x, \alpha)$
    return $\mathrm{K}(x)\alpha$
end function

function $\mathrm{F}(x, \alpha)$
    return $-\frac{1}{2}\nabla_x \left(\alpha^T \mathrm{K}(x)\alpha\right)$ ▷ in closed form or by finite differences
end function

```
function K(x)
    return K(x) (or G(x)^{-1})                          ▷ in closed form
end function
```

Appendix B. Proofs.

B.1. Transport and connection. We prove a result connecting successive covariant derivatives to parallel transport:

Proposition B.1. *Let V be a vector field on $\mathcal{M}$. Let $\gamma : [0, 1] \to \mathcal{M}$ be a geodesic. Then:*

$$\nabla^k_{\dot\gamma} V(\gamma(t)) = \frac{\mathrm{d}^k}{\mathrm{d}h^k} P^{-1}_{t,t+h}(V(\gamma(t+h)) \tag{18}$$

Proof. Let $E_i(0)$ be an orthonormal basis of $T_{\gamma(0)}\mathcal{M}$. Using the parallel transport along γ, we get orthonormal basis $E_i(s)$ of $T_{\gamma(t)}\mathcal{M}$ for all t. We have:

$$\frac{\mathrm{d}^k}{\mathrm{d}h^k} P^{-1}_{t,t+h}(V(\gamma(t+h)) = \frac{\mathrm{d}^k}{\mathrm{d}h^k} P^{-1}_{t,t+h} \sum_{i=1}^n a_i(t+h)E_i(t+h) = \sum_{i=1}^n \frac{\mathrm{d}^k a_i(t+h)}{\mathrm{d}h^k} E_i(t).$$

On the other hand:

$$\nabla^k_{\dot\gamma} V(\gamma(t)) = \nabla^k_{\dot\gamma} \sum_{i=1}^n a_i(t)E_i(t) = \sum_{i=1}^n \nabla^k_{\dot\gamma}(a_i(t))E_i(t) = \sum_{i=1}^n \frac{\mathrm{d}^k a_i(t+h)}{\mathrm{d}h^k} E_i(t)$$

by definition of $E_i(s)$. ∎

B.2. Proof that we can compute the geodesic simultaneously with a second-order method. We give here a control on the error made in the scheme when computing the main geodesic approximately and simultaneously with the parallel transport. We assume that the main geodesic is computed with a second-order method, and we need to control the subsequent error on the Jacobi field. The computations are made in coordinates, and the error measured by the 2-norm on those coordinates.

Proposition B.2. *There exists $A > 0$ such that for all $t \in [0, 1[$, for all $h \in [0, 1-t]$, for all $w \in T_{\gamma(t)}\mathcal{M}$:*

$$\left\| \frac{\mathrm{J}^{\tilde w_k}_{\gamma_k}(h)}{h} - \frac{\mathrm{J}^{\tilde w_k}_{\tilde\gamma_k}(h)}{h} \right\|_2 \leq Ah^2$$

Proof. Let $t \in [0, 1[$, for all $h \in [0, 1-t]$, for all $w \in T_{\gamma(t)}\mathcal{M}$. As previouslt, the term rewrites :

$$\left\| \frac{\mathrm{J}^{\tilde w_k}_{\gamma_k}(h)}{h} - \frac{\mathrm{J}^{\tilde w_k}_{\tilde\gamma_k}(h)}{h} \right\|_2 = \left\| \frac{\partial \mathrm{Exp}_\gamma(h\dot\gamma + xw)}{\partial x}\bigg|_{x=0} - \frac{\partial \mathrm{Exp}_{\tilde\gamma}(h\dot{\tilde\gamma} + xw)}{\partial x}\bigg|_{x=0} \right\|_2 \tag{19}$$

This is the difference between the derivatives of two solutions of the same differential equation (5) with respect to an initial parameter. More precisely, we define Π : $\Phi(K) \times B_{\mathbb{R}^n}(0, \|\tilde\gamma_k\| + 2\varepsilon\|\tilde w_k\|) \times [0, \eta]) \to \mathbb{R}^n$ such that $\Pi(p_0, \alpha_0, h)$ are the coordinates of the solutions of the Hamiltonian equation at time h with initial coordinates p_0 and initial velocity α_0. Π is the flow, in coordinates, of the geodesic equation. We can now rewrite Equation (19):

$$\left\| \frac{J^{\tilde{w}_k}_{\gamma_k}(h)}{h} - \frac{J^{\tilde{w}_k}_{\tilde{\gamma}_k}(h)}{h} \right\|_2 = \left\| \frac{\partial \Pi(\gamma_k, \dot{\gamma}_k + \varepsilon \tilde{w}_k, h)}{\partial \varepsilon}\Big|_{\varepsilon=0} - \frac{\partial \Pi(\tilde{\gamma}_k, \dot{\tilde{\gamma}}_k + \varepsilon \tilde{w}_k, h)}{\partial \varepsilon}\Big|_{\varepsilon=0} \right\|_2$$

By Cauchy-Lipschitz theorem, the flow Π of the Hamiltonian equation is smooth. Hence, its derivatives are bounded over its compact set of definition. Hence there exists a constant A such that:

$$\left\| \frac{J^{\tilde{w}_k}_{\gamma_k}(h)}{h} - \frac{J^{\tilde{w}_k}_{\tilde{\gamma}_k}(h)}{h} \right\|_2 \leq A \left(\|\tilde{\gamma} - \gamma\|_2 + \|\dot{\tilde{\gamma}} - \dot{\gamma}\|_2 \right)$$

where we can once again assume A independent of t or h. In coordinates, we use a second-order Runge-Kutta method to integrate the geodesic equation so that the cumulated error is of order h^2. Hence, there exists a positive constant B which does not depend on h, t or w such that :

$$\left\| \frac{J^{\tilde{w}_k}_{\gamma_k}(h)}{h} - \frac{J^{\tilde{w}_k}_{\tilde{\gamma}_k}(h)}{h} \right\|_2 \leq Bh^2.$$

□

B.3. Numerical approximation with a single perturbed geodesic. We suppose here that the computation to get the Jacobi field is done with a first-order method i.e. with the computation of a single perturbed geodesic computed with a second-order Runge-Kutta method. We prove the following lemma :

LEMMA B.3. *For all $L > 0$, There exists $A > 0$ such that for all $t \in [0, 1[$, for all $h \in [0, 1-t]$, for all $w \in T_{\gamma(t)}\mathcal{M}$ with $\|w\|_2 < L$ –in the global system of coordinates – we have:*

$$\left\| \frac{J^w_{\gamma(t)}(h) - \tilde{J}^w_{\gamma(t)}(h)}{h} \right\|_2 \leq A(h^2 + \varepsilon h)$$

where $\tilde{J}^w_{\gamma(t)}(h)$ is the numerical approximation of $J^w_{\gamma(t)}(h)$ computed with a single perturbed geodesic and a first-order differentiation method. We consider that this approximation is computed in the global system of coordinates.

Proof. Let $L > 0$. Let $t \in [0, 1[$, $h \in [0, 1-t]$ and $w \in T_{\gamma(t)}\mathcal{M}$. We split the error term in two parts :

$$\left\| \frac{J^w_{\gamma(t)}(h)}{h} - \frac{\tilde{J}^w_{\gamma(t)}(h)}{h} \right\|_2 \leq \underbrace{\left\| \frac{J^w_{\gamma(t)}(h)}{h} - \frac{\mathrm{Exp}[h(\dot{\gamma}(t) + \varepsilon w)] - \mathrm{Exp}[h\dot{\gamma}(t)]}{\varepsilon h} \right\|_2}_{(1)}$$
$$+ \underbrace{\left\| \frac{\mathrm{Exp}[(h(\dot{\gamma}(t) + \varepsilon w)] - \mathrm{Exp}[h\dot{\gamma}(t)]}{\varepsilon h} - \frac{\tilde{\mathrm{Exp}}[h(\dot{\gamma}(t) + \varepsilon w)] - \tilde{\mathrm{Exp}}[h\dot{\gamma}(t)]}{\varepsilon h} \right\|_2}_{(2)}$$

where Exp is the Riemannian exponential at $\gamma(t)$ and $\tilde{\mathrm{Exp}}$ is the numerical approximation of this Riemannian exponential computed thanks to the Hamiltonian equations. When running the scheme, these computations are done in the global system of coordinates.

(1). Let $i \in \{1, ..., n\}$ and let $F^i : (x, t, w) \to \mathrm{Exp}[h\dot{\gamma}(t) + xw]^i$. We have:

$$\begin{aligned}\frac{\mathrm{J}^w_{\gamma(t)}(h)^i}{h} &- \frac{\mathrm{Exp}[h(\dot{\gamma}(t)+\varepsilon w)]^i - \mathrm{Exp}[h\dot{\gamma}(t)]^i}{\varepsilon h} \\ &= \frac{1}{h}\frac{\partial F^i(\varepsilon h, t, w)}{\partial \varepsilon} - \frac{F^i(\varepsilon h, t, w) - F^i(0, w)}{\varepsilon h} \\ &= \left.\frac{\partial F^i(x,t,w)}{\partial x}\right|_{x=0} - \frac{F^i(\varepsilon h, t, w) - F^i(0, t, w)}{\varepsilon h}\end{aligned}$$

Now, F^i is smooth hence its derivatives are bounded over the compact set $[0, \eta] \times [0, 1] \times B_{\mathbb{R}^n}(0, L)$. Using the mean-value theorem, there exists $B > 0$ such that for all i, for all t, for all h and for all w with $\|w\|_2 \leq L$:

$$\left| \frac{\mathrm{J}^w_{\gamma(t)}(h)^i}{h} - \frac{\mathrm{Exp}[h\dot{\gamma}(t) + \varepsilon h w]^i - \mathrm{Exp}[h\dot{\gamma}(t)]^i}{\varepsilon h} \right| \leq B\varepsilon h$$

so that there exists $C > 0$ such that for all t, for all h and for all w with $\|w\|_2 \leq L$:

$$\left\| \frac{\mathrm{J}^w_{\gamma(t)}(h)}{h} - \frac{\mathrm{Exp}[h\dot{\gamma}(t) + \varepsilon h w] - \mathrm{Exp}[h\dot{\gamma}(t)]}{\varepsilon h} \right\|_2 \leq C\varepsilon h$$

(2). We rewrite the Hamiltonian equation $\dot{x}(t) = F_1(x(t), \alpha(t))$ and $\dot{\alpha}(t) = F_2(x(t), \alpha(t))$. We note $x_\varepsilon, \alpha_0^\varepsilon$ the solution of this equation (in the global system of coordinates) with initial conditions $x_\varepsilon(0) = x_0$ and $\alpha_0^\varepsilon = K(x_0)^{-1}(\dot{\gamma} + \varepsilon w)$. The term (2) rewrites:

$$\tfrac{1}{\epsilon h}\|(x^\varepsilon(h) - x^0(h)) - (\tilde{x}^\varepsilon(h) - \tilde{x}^0(h))\|_2$$

First, we develop x^ε in the neighborhood of 0:

$$x^\varepsilon(h) = x_0 + h\dot{x}^\varepsilon(0) + \frac{h^2}{2}\ddot{x}^\varepsilon(0) + \int_0^h \frac{(h-t)^2}{2}\dddot{x}^\varepsilon(t)\mathrm{d}t \tag{20}$$

We have, for the last term:

$$\left\| \int_0^h \frac{(h-t)^2}{2}\dddot{x}^\varepsilon(t)\mathrm{d}t - \int_0^h \frac{(h-t)^2}{2}\dddot{x}^0(t)\mathrm{d}t \right\|_2 = \left\| \int_0^h \int_0^{+\varepsilon} \frac{(h-t)^2}{2}\partial_\varepsilon \dddot{x}^\varepsilon(u, t)\mathrm{d}t\mathrm{d}u \right\|_2$$

x^ε being solution of a smooth ordinary differential equation with smoothly varying initial conditions, it is smooth in time and with respect to ε. Hence, when the initial conditions are within a compact, $\partial_\varepsilon \dddot{x}^\varepsilon$ is bounded, hence there exists $D > 0$ such that:

$$\left\| \int_0^h \frac{(h-t)^2}{2}\dddot{x}^\varepsilon(t)\mathrm{d}t - \int_0^h \frac{(h-t)^2}{2}\dddot{x}^0(t)\mathrm{d}t \right\|_2 \leq Dh^3\varepsilon$$

For the other terms:

$$\dot{x}^\varepsilon(0) = K(x_0)\alpha_0 = \dot{\gamma} + \varepsilon w$$

and

$$\ddot{x}^\varepsilon(0) = \left.\frac{\mathrm{d}K(x^\varepsilon(t))\alpha^\varepsilon(t)}{\mathrm{d}t}\right|_{t=0} = (\nabla_x K)(x_0)(\dot\gamma + \varepsilon w)\alpha_0^\varepsilon + K(x_0)F_2(x_0, \alpha_0^\varepsilon)$$

Now we focus on the approximation that we compute with the second-order Runge-Kutta scheme, denoting it with a tilde:

$$\tilde{x}^\varepsilon(h) = x_0 + hF_1\left(x_0 + \tfrac{h}{2}F_1(x_0, \alpha_0^\varepsilon), \alpha_0^\varepsilon + \tfrac{h}{2}F_2(x_0, \alpha_0^\varepsilon)\right)$$

We replace F_1 and α_0^ε by their expressions:

$$\begin{aligned}\tilde{x}^\varepsilon(h) &= x_0 + hK\left(x_0 + \tfrac{h}{2}F_1(x_0, \alpha_0^\varepsilon)\right)\left(\alpha_0^\varepsilon + \tfrac{h}{2}F_2(x_0, \alpha_0^\varepsilon)\right)\\ &= x_0 + hK\left(x_0 + \tfrac{h}{2}(\dot\gamma + \varepsilon w)\right)\left(\alpha_0^\varepsilon + \tfrac{h}{2}F_2(x_0, \alpha_0^\varepsilon)\right)\end{aligned}$$

We use a Taylor expansion for K:

$$K(x_0 + \tfrac{h}{2}(\dot\gamma + \varepsilon w)) = K(x_0) + \tfrac{h}{2}(\nabla_x K)(x_0)(\dot\gamma + \varepsilon w) + \tfrac{h^2}{8}\nabla^2 K(x_0)(\dot\gamma + \varepsilon w) + \mathrm{O}\big(h^3\big)$$

So that:

$$\begin{aligned}\tilde{x}^\varepsilon(h) &= x_0 + h(\dot\gamma + \varepsilon w) + \tfrac{h^2}{2}\Big[K(x_0)F_2(x_0, \alpha_0^\varepsilon) + \nabla_x K(x_0)(\dot\gamma + \varepsilon w)\alpha_0^\varepsilon\Big]\\ &+ \tfrac{h^3}{4}\Big[\nabla_x K(x_0)(\dot\gamma + \varepsilon w)F_2(x_0, \alpha_0^\varepsilon) + \tfrac{1}{2}\nabla^2 K(x_0)(\dot\gamma + \varepsilon w)\alpha_0^\varepsilon\Big] + O(h^4)\end{aligned}$$

The third order terms of $x^\varepsilon - x^0$ is:

$$\begin{aligned}&\nabla_x K(x_0)\Big[(\dot\gamma + \varepsilon w)F_2(x_0, \alpha_0^\varepsilon) - (\dot\gamma)F_2(x_0, \alpha_0^0)\Big]\\ &+ \tfrac{1}{2}\Big[\nabla^2 K(x_0)(\dot\gamma + \varepsilon w)\alpha_0^\varepsilon - \nabla^2 K(x_0)(\dot\gamma)\alpha_0^0\Big]\end{aligned}$$

Both these terms are the differences of smooth functions at points whose distance is of order $\varepsilon\|w\|_2$. Because those functions are smooth, and we are only interested in these majorations for points in K and tangent vectors in a compact ball in the tangent space, this third order term is bounded by $Eh^3\varepsilon\|w\|_g$ where E is a positive constant which does not depend on the position on the geodesic. Finally, the differences between the second order terms of x^ε and $\tilde{x}^\varepsilon$ is zero, so that :

$$\|(x^\varepsilon(h) - x^0(h)) - (\tilde{x}^\varepsilon(h) - \tilde{x}^0(h))\|_2 \leq (Dh^3\varepsilon + Eh^3\varepsilon)\|w\|_g$$

which concludes. ◻

B.4. Numerical approximation with two perturbed geodesics. We suppose here that the computation to get the Jacobi field is done with a central finite difference method. We prove the following lemma:

LEMMA B.4. *For all $L > 0$, there exists $A > 0$ such that for all $t \in [0, 1[$, for all $h \in [0, 1 - t]$, for all $w \in T_{\gamma(t)}\mathcal{M}$ with $\|w\|_2 < L$ –in the global system of coordinates – we have:*

$$\left\|\frac{\mathrm{J}^w_{\gamma(t)}(h) - \tilde{\mathrm{J}}^w_{\gamma(t)}(h)}{h}\right\|_2 \leq A(h^2 + \varepsilon h)$$

where $\tilde{\mathrm{J}}^w_{\gamma(t)}(h)$ *is the numerical approximation of* $\mathrm{J}^w_{\gamma(t)}(h)$ *computed with two perturbed geodesics and a central finite differentiation method. We consider that this approximation is computed in the global system of coordinates.*

The proof is similar to the one above.

REFERENCES

[1] Kheyfets, A., Miller, W.A., Newton, G.A.: Schild's ladder parallel transport procedure for an arbitrary connection. International Journal of Theoretical Physics **39**(12), 2891–2898 (2000)

[2] Lenglet, C., Rousson, M., Deriche, R., Faugeras, O.: Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor mri processing. Journal of Mathematical Imaging and Vision **25**(3), 423–444 (2006). DOI 10.1007/s10851-006-6897-z. URL http://dx.doi.org/10.1007/s10851-006-6897-z

[3] Lorenzi, M., Ayache, N., Pennec, X.: Schild's ladder for the parallel transport of deformations in time series of images. In: Biennial International Conference on Information Processing in Medical Imaging, pp. 463–474. Springer (2011)

[4] Lorenzi, M., Pennec, X.: Geodesics, Parallel Transport & One-parameter Subgroups for Diffeomorphic Image Registration. International Journal of Computer Vision **105**(2), 111–127 (2013). DOI 10.1007/s11263-012-0598-4. URL https://hal.inria.fr/hal-00813835

[5] Lorenzi, M., Pennec, X.: Parallel Transport with Pole Ladder: Application to Deformations of time Series of Images. In: F. Nielsen, F. Barbaresco (eds.) GSI2013 - Geometric Science of Information, vol. 8085, pp. 68–75. Springer, Paris, France (2013). DOI 10.1007/978-3-642-40020-9_6. URL https://hal.inria.fr/hal-00819898

[6] Manfredo, P.: Riemannian geometry (1992)

[7] Schiratti, J.B.: Methods and algorithms to learn spatio-temporal changes from longitudinal manifold-valued observations. PhD Thesis (2017)

[8] Schiratti, J.B., Allassonnière, S., Colliot, O., Durrleman, S.: Learning spatiotemporal trajectories from manifold-valued longitudinal data. In: NIPS 28

[9] Vogtmann, K., Weinstein, A., Arnol'd, V.: Mathematical Methods of Classical Mechanics. Graduate Texts in Mathematics. Springer New York (1997). URL https://books.google.fr/books?id=Pd8-s6rOt_cC

[10] Younes, L.: Jacobi fields in groups of diffeomorphisms and applications. Quarterly of Applied Mathematics **65**(1), 113–134 (2007). URL http://www.jstor.org/stable/43638762

[11] Younes, L.: Shapes and diffeomorphisms. Heidelberg : Springer (2010). "A direct application of what is presented in the book is a branch of the computerized analysis of medical images called computational anatomy"–Back cover

[12] Zhang, M., Fletcher, P.: Probabilistic principal geodesic analysis. In: C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K.Q. Weinberger (eds.) Advances in Neural Information Processing Systems 26, pp. 1178–1186. Curran Associates, Inc. (2013). URL http://papers.nips.cc/paper/5133-probabilistic-principal-geodesic-analysis.pdf