SIAM J. NUMER. ANAL.
Vol. 56, No. 4, pp. 2563–2584

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

MAXIME LOUIS†, BENJAMIN CHARLIER‡, PAUL JUSSELIN§, SUSOVAN PAL†, AND STANLEY DURRLEMAN†

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 for computing parallel transport require either the computation of Riemannian logarithms, such as Schild's ladder, or the Christoffel symbols. The logarithm is rarely given in closed form, and therefore expensive to compute, whereas the Christoffel symbols are in general hard and costly to compute. From an identity between parallel transport and Jacobi fields, we propose a numerical scheme to approximate parallel transport along a geodesic. We find and prove an optimal convergence rate for the scheme, which is equivalent to Schild's ladders. We investigate potential variations of the scheme and give experimental results on the Euclidean 2-sphere and on the manifold of symmetric positive definite matrices.

Key words. parallel transport, Riemannian manifold, numerical scheme, Jacobi field

AMS subject classifications. 53C21, 65L12, 65L06, 34G10

DOI. 10.1137/17M1130617

1. Introduction.

Riemannian geometry has long been contained within the field of pure mathematics and theoretical physics. Nevertheless, there is an emerging trend to use the tools of 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 positive definite matrices, diffeomorphisms, or bounded measurements. Such data may be considered therefore as points on a Riemannian manifold and analyzed by specific statistical approaches [14, 4, 10, 5]. At the core of these approaches lies parallel transport, an isometry between tangent spaces 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 efficient numerical schemes to compute parallel transport on manifolds.

*Received by the editors May 16, 2017; accepted for publication (in revised form) March 21, 2018; published electronically August 9, 2018.

http://www.siam.org/journals/sinum/56-4/M113061.html

Funding: This work was partially funded by the European Research Council (ERC) under grant agreement 678304, the European Union's Horizon 2020 research and innovation programme under grant agreement 666992, and the program "Investissements d'avenir" ANR-10-IAIHU-06.

†Inria Paris, Aramis Project-Team, 75013 Paris, France, and Sorbonne Universités, UPMC Univ. Paris 06, Inserm, CNRS, Institut du Cerveau et de la Moelle (ICM) - Hôpital Pitié-Salpêtrière, F-75013 Paris, France (maxime.louis.x2012@gmail.com, susovan97@gmail.com, stanley.durrleman@inria.fr).

‡Institut Montpelliérain Alexander Grothendieck, CNRS, Université de Montpellier, Montpellier 34090, France, Inria Paris, Aramis Project-Team, 75013 Paris, France, and Sorbonne Universités, UPMC Univ. Paris 06, Inserm, CNRS, Institut du Cerveau et de la Moelle (ICM) - Hôpital Pitié-Salpêtrière, F-75013 Paris, France (benjamin.charlier@umontpellier.fr).

§CMLA, ENS Cachan, Cachan Ile-De-France 94230, France, Inria Paris, Aramis Project-Team, 75013 Paris, France, and Sorbonne Universités, UPMC Univ. Paris 06, Inserm, CNRS, Institut du Cerveau et de la Moelle (ICM) - Hôpital Pitié-Salpêtrière, F-75013 Paris, France (pjusseli@ens-cachan.fr).

The parallel transport of a given tangent vector is defined as the solution of an ODE (see [2, page 52]), written in terms of the Christoffel symbols. The computation of the Christoffel symbols requires access to the metric coefficients and their derivatives, making the equation integration using standard numerical schemes very costly in situations where no closed-form formulas are available for the metric coefficients or their derivatives.

An alternative is to use Schild's ladder [3], or its faster version in the case of geodesics, the pole ladder [7]. These schemes essentially require the computation of Riemannian exponentials and logarithms at each step. Usually, the computation of the exponential may be done by integrating Hamiltonian equations and does not raise specific difficulties. By contrast, the computation of the logarithm must often be done by solving an inverse problem with the use of an optimization scheme such as a gradient descent. Such optimization schemes are approximate and sensitive to the initial conditions and to hyperparameters, which leads to additional numerical errors—most of the time uncontrolled—as well as an increased computational cost. 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–Hausdorff formula (see [6]), Schild's ladder is an efficient alternative. When this is not the case, it becomes hardly tractable. A more detailed analysis of the convergence of Schild's ladder method can be found in [9].

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 a second-order error. This idea has been suggested in [12] with further credits to [1], but without either a formal definition or 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 accuracy. We will show how to use only the inverse of the metric and its derivatives when performing the different steps of the scheme. This different set of requirements makes the scheme attractive in a set of situations different from the integration of the ODE or Schild's ladder. 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 Schild's ladder in practice. Thus, they will show that this scheme offers a compelling alternative to computing parallel transport 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 times $t > 0$ on a smooth manifold $\mathcal{M}$ of finite dimension $n \in \mathbb{N}$ provided with a smooth Riemannian metric g. We denote the Riemannian exponential as Exp and ∇ as the covariant derivative. For $p \in \mathcal{M}$, $T_p\mathcal{M}$ denotes the tangent space of $\mathcal{M}$ at p. For all $s, t \geq 0$ and for all $w \in T_{\gamma(s)}\mathcal{M}$, we denote by $\mathrm{P}_{s,t}(w) \in T_{\gamma(t)}\mathcal{M}$ 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 denote by $\mathrm{J}^w_{\gamma(t)}(h)$

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, [2, pages 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 by $\|\cdot\|_g$ the Riemannian norm on the tangent spaces defined from the metric g and by $g_p : T_p\mathcal{M} \times T_p\mathcal{M} \to \mathbb{R}$ the metric at any $p \in \mathcal{M}$. We use Einstein notations.

We fix Ω a compact subset of $\mathcal{M}$ such that Ω contains a neighborhood of $\gamma([0,1])$. We also set $w \in T_{\gamma(0)}\gamma$ and $w(t) = P_{0,t}(w)$. We suppose that there exists a coordinate system on Ω, and we denote by $\Phi : \Omega \longrightarrow U$ 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 of Ω; we denote by $\left.\frac{\partial}{\partial x^i}\right|_p$ the ith element of the corresponding basis of $T_p\mathcal{M}$ for any $p \in \mathcal{M}$. Note finally that, since the injectivity radius is a smooth function of the position on the manifold (see [2]) and since it is everywhere positive on Ω, there exists $\eta > 0$ such that for all p in Ω, the injectivity radius at p is larger than η.

The problem in this paper is to provide a way to compute an approximation of $P_{0,1}(w)$.

We suppose throughout the paper the existence of a single coordinate chart defined on Ω. In this setting, we propose a numerical scheme which gives an error varying linearly with the size of the integration step. Once this result is established, since in any case $\gamma([0,1])$ can be covered by finitely many charts, it is possible to apply the proposed method to parallel transport on each chart successively. The errors during this computation of the parallel transport would increase, but the convergence result remains valid.

2.2. The key identity. The numerical scheme that we propose arises from the following identity, which is mentioned in [12]. 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 $X(0) = w$, where $(\Gamma^i_{kl})_{i,j,k\in\{1,\dots,n\}}$ are the Christoffel symbols associated with the Levi-Civita connection for the metric g. 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 reads as

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

By definition, the ith normal coordinate of $\mathrm{Exp}_{\gamma(0)}\,(t(v_0 + \varepsilon w))$ is $t(v_0^i + \varepsilon w^i)$. Therefore, the ith 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. ∎

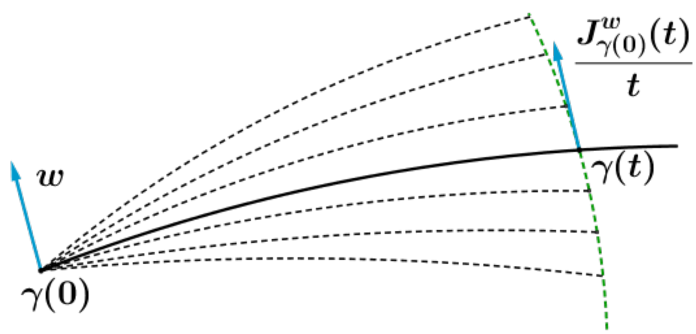

FIG. 1. *The solid line is the geodesic. The green dotted line is formed by the perturbed geodesics at time t. The blue arrows are the initial vector and its approximated parallel transport at time t.*

This control on the approximation of the transport by a Jacobi field suggests dividing $[0,1]$ into N intervals $[\frac{k}{N}, \frac{k+1}{N}]$ of length $h = \frac{1}{N}$ for $k = 0, \ldots, N-1$ and approximating 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 a 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 (2) alone raises a numerical error 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 $\mathbb{R}^3$. Using spherical coordinates (θ, ϕ) on the sphere, chosen so that the whole geodesic is in the coordinate chart, we get coordinates on the tangent space at any point $\gamma(t)$. In this spherical system of coordinates, it is straightforward to see that the parallel transport of $w = p \times v$ along $\gamma(t)$ has constant coordinates, where $\times$ denote the usual cross-product on $\mathbb{R}^3$.

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,

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

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

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

yielding

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

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

3. The numerical scheme.

3.1. The algorithm. In general, there are no closed-form 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. Let $x(t) = (x_1(t),\dots,x_d(t))^T$ be the coordinates of $\gamma(t)$ in a given local chart, and let $\alpha(t) = (\alpha_1(t),\dots,\alpha_d(t))^T$ be the coordinates of the momentum $g_{\gamma(t)}(\dot\gamma(t),\cdot) \in T^*_{\gamma(t)}\mathcal{M}$ in the same local chart. We have then (see [13])

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

where $K(x(t))$, a d-by-d matrix, is the inverse of the metric g expressed in the local chart. Note that using (5) to integrate the geodesic equation will require us to convert initial tangent vectors into initial momenta, as seen in the algorithm description below.

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

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

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

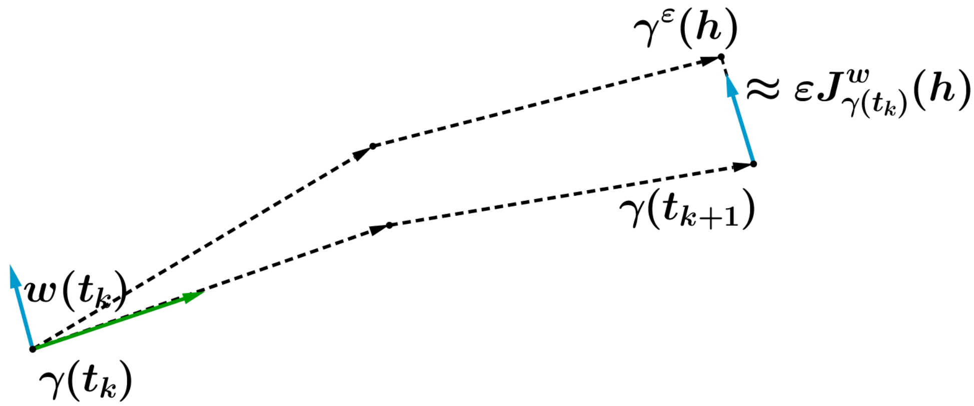

FIG. 2. *One step of the numerical scheme. The dotted arrows represent the steps of the Runge–Kutta integrations for the main geodesic* γ *and for the perturbed geodesic* γ^ε*. The blue arrows are the initial* $w(t_k)$ *and the obtained approximated transport using* (6) *with* $h = t_{k+1} - t_k$.

The algorithm. Let $N \in \mathbb{N}$. We divide $[0,1]$ into N intervals $[t_k, t_{k+1}]$ with $t_k = \frac{k}{N}$ and denote by $h = \frac{1}{N}$ the size of the integration step. We initialize $\gamma_0 = \gamma(0)$, $\dot{\gamma}_0 = \dot{\gamma}(0)$, $\tilde{w}_0 = w$ and solve $\tilde{\beta}_0 = K^{-1}(\gamma_0)\tilde{w}_0$ and $\tilde{\alpha}_0 = K^{-1}(\gamma_0)\dot{\gamma}_0$. We propose to compute, at step k, the following:

(i) the new point $\tilde{\gamma}_{k+1}$ and momentum $\tilde{\alpha}_{k+1}$ of the main geodesic by performing one step of length h of a second-order Runge–Kutta method on (5);

(ii) the perturbed geodesic starting at $\tilde{\gamma}_k$ with initial momentum $\tilde{\alpha}_k + \varepsilon\tilde{\beta}_k$ at time h, which we denote by $\tilde{\gamma}^\varepsilon_{k+1}$, by performing one step of length h of a second-order Runge–Kutta method on (5);

(iii) the estimated parallel transport

$$\hat{w}_{k+1} = \frac{\tilde{\gamma}^\varepsilon_{k+1} - \tilde{\gamma}_{k+1}}{h\varepsilon}; \tag{7}$$

(iv) the corresponding momentum $\hat{\beta}_{k+1}$ by solving $K(\tilde{\gamma}_{k+1})\hat{\beta}_{k+1} = \hat{w}_{k+1}$.

At the end of the scheme, $\tilde{w}_N$ is the proposed approximation of $P_{0,1}(w)$. Figure 2 illustrates the principle. A complete pseudocode 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 step (iv) of the algorithm requires solving a linear system of size n. Solving the linear system can be done with a complexity less than cubic in the dimension (in $\mathrm{O}\big(n^{2.374}\big)$ using the Coppersmith–Winograd algorithm).

3.2. Possible variations. There are a few possible variations of the presented algorithm.

1. The first variation is to use higher-order Runge–Kutta methods to integrate the geodesic equations at steps (i) and (ii). We prove that a second-order integration of the geodesic equation is enough to guarantee convergence and notice experimentally the absence of convergence with a first-order integration of the geodesic equation. Experiments indicate a linear convergence with an

improved constant using this variation. Depending on the situation, the extra computations required at each step may be counterbalanced by this increased precision.

2. The second variation uses a higher-order finite difference scheme by replacing steps (ii) and (iii) in the following way. At the kth iteration, compute two perturbed geodesics starting at $\tilde{\gamma}_k$ and with initial momentum $\tilde{\alpha}_k + \epsilon\tilde{\beta}_k$ (resp., $\tilde{\alpha}_k - \epsilon\tilde{\beta}_k$) at time h, which we denote by $\tilde{\gamma}_{k+1}^{+\varepsilon}$ (resp., $\tilde{\gamma}_{k+1}^{-\varepsilon}$), by performing one step of length h of a second-order Runge–Kutta method on (5). Then proceed to a second-order differentiation to approximate the Jacobi field, and set

$$\hat{w}_{k+1} = \frac{\tilde{\gamma}_{k+1}^{+\varepsilon} - \tilde{\gamma}_{k+1}^{-\varepsilon}}{2h\varepsilon}. \tag{8}$$

Empirically, this variation does not seem to bring any substantial improvement to the scheme.

3. The final variation of the scheme consists in adding an extra renormalization step at the end of each iteration:

(v) Renormalize the momentum and the corresponding vector using

$$\tilde{\beta}_{k+1} = a_k\hat{\beta}_{k+1} + b_k\tilde{\alpha}_{k+1},$$

$$\tilde{w}_{k+1} = K(\tilde{\gamma}_{k+1})\tilde{\beta}_{k+1},$$

where a_k and b_k are factors ensuring $\tilde{\beta}_{k+1}^\top K(\tilde{\gamma}_{k+1})\tilde{\beta}_{k+1} = \beta_0^\top K(\gamma_0)\beta_0$ and $\tilde{\beta}_{k+1}^\top K(\tilde{\gamma}_{k+1})\tilde{\alpha}_{k+1} = \beta_0^\top K(\gamma_0)\alpha_0$. Indeed, the quantities $\beta(t)^\top K(\gamma(t))\beta(t)$ and $\beta(t)^\top K(\gamma(t))\alpha(t)$ are preserved along the parallel transport. This extra step is cheap even when the dimension is large. Empirically, it leads to the same rate of convergence with a smaller constant.

We will show that the proposed algorithm and variations 1 and 2 ensure convergence of the final estimate. We do not prove convergence with variation 3, but this additional step can be expected to improve the quality of the approximation at each step, at least when the discretization is sufficiently thin, by enforcing the conversation of quantities which should be conserved. Note that the best accuracy for a given computational cost is not necessarily obtained with the method in section 3.1, but might be attained with one of the proposed variations, as a few more computations at each step may be counterbalanced by a smaller constant in the convergence rate.

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

THEOREM 3.1. *We suppose here the hypotheses stated in section* 2.1. *Let $N \in \mathbb{N}$ be the number of integration steps. Let $w \in T_{\gamma(0)}\mathcal{M}$ be the vector to be transported. We denote the error*

$$\delta_k = \|P_{0,t_k}(w) - \tilde{w}_k\|_2,$$

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 the chart Φ on Ω. We denote by ε the parameter used in step* (ii) *and $h = \frac{1}{N}$ the size of the step used for the Runge–Kutta approximate solution of the geodesic equation. If we take $\varepsilon = h$, then we have*

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

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. Further decreasing ε has no visible effect on the accuracy of the estimation, and 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 $\gamma(1)$ is not accessible, but only is its approximation $\tilde{\gamma}_N$ computed by the Runge–Kutta integration of the Hamiltonian equation. However, Theorem 3.1 implies that the couple $(\tilde{\gamma}_N, \tilde{w}_N)$ converges towards $(\gamma(1), P_{0,1}(w))$ using the ℓ^2 distance on $\mathcal{M} \times T\mathcal{M}$ and a coordinate system in a neighborhood of $\gamma(1)$, which is equivalent to any distance on $\mathcal{M} \times T\mathcal{M}$ on this neighborhood and hence is the right notion of convergence.

We give the proof in the next section. The technical lemmas used in the proof are all in Appendix B: in Appendix B.1, we prove an intermediate result allowing uniform controls on norms of tensors; in Appendix B.3, we prove a stronger result than Proposition 2.1 with stronger hypotheses; and in Appendix B.4, we prove a result allowing us to control the accumulation of the error.

4. Proof of Theorem 3.1. We prove the convergence of the algorithm.

Proof. We will denote, as in the description of the algorithm in section 3, by $\gamma_k = \gamma(t_k)$, $\tilde{\gamma}_k = \tilde{\gamma}(t_k)$ its approximation in the algorithm. Let N be a number of discretization steps, and let $k \in \{1, \ldots, N\}$. 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

- $\tilde{\gamma}_k$ is the approximation of the geodesic coordinates at step k,
- $w_k = w(t_k)$ 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: $\tilde{\mathrm{J}}^{\tilde{w}_k}_{\tilde{\gamma}_k} = \frac{\tilde{\gamma}^{\varepsilon}_{k+1} - \tilde{\gamma}_{k+1}}{\varepsilon}$, and
- $\mathrm{J}^{\tilde{w}_k}_{\tilde{\gamma}_k}(h)$ is the exact Jacobi field computed with the approximations $\tilde{w}$, $\tilde{\gamma}$, and $\tilde{\dot{\gamma}}$ i.e., the Jacobi field defined from the geodesic with initial position $\tilde{\gamma}_k$ and initial momentum $\tilde{\alpha}_k$, with a perturbation $\tilde{w}_k$.

We provide upper bounds for each of these terms. We start by assuming $\|w_k\|_2 \leq 2\|w_0\|_2$, before showing that it is verified for any $k \leq N$ when N is large enough. We could assume more generally that$\|w_k\|_2 \leq C\|w_0\|_2$ for any $C > 1$. The idea is to get a uniform control on the errors at each step by assuming that $\|w_k\|_2$ does not grow too much, and to show afterwards that the control we get is tight enough to ensure, when the number of integration steps is large, that we do have $\|w_k\|_2 \leq 2\|w_0\|_2$.

Term (1). This is the intrinsic error when using the Jacobi field. We show in Proposition B.3 that for h small enough

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

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

$$\left\| P_{t_k,t_{k+1}}(w_k) - \frac{\mathrm{J}^{w_k}_{\gamma(k)}(h)}{h} \right\|_2 \leq A'h^2\|w_k\|_2 \leq 2A'h^2\|w_0\|_2. \tag{9}$$

Term (2). Lemma B.4 shows that for h small enough

$$\left\| \frac{\mathrm{J}^{w_k}_{\gamma(t_k)}(h)}{h} - \frac{\mathrm{J}^{\tilde{w}_k}_{\gamma(t_k)}(h)}{h} \right\|_2 \leq (1+Bh)\delta_k. \tag{10}$$

Term (3). This term measures the error linked to our approximate knowledge of the geodesic γ. It is proved in Appendix B.5 that there exists a constant $C > 0$ which does not depend on k or h such that

$$\left\| \frac{\mathrm{J}^{\tilde{w}_k}_{\gamma_k}(h)}{h} - \frac{\tilde{\mathrm{J}}^{\tilde{w}_k}_{\gamma_k}(h)}{h} \right\|_2 \leq Ch^2. \tag{11}$$

Term (4). This is the difference between the analytical computation of J and its approximation. It is proved in Appendices B.6 and B.7 that if we use a Runge–Kutta method of order 2 to compute the geodesic points $\gamma^{\varepsilon}_{k+1}$ and γ_{k+1} and a first-order differentiation to compute the Jacobi field as described in step (iii) of the algorithm, or if we use two perturbed geodesics $\gamma^{\varepsilon}_{k+1}$ and $\gamma^{-\varepsilon}_{k+1}$ and a second-order differentiation method to compute the Jacobi field as described in (8), there exists $D \geq 0$ which does not depend on k such that

$$\left\| \frac{\mathrm{J}^{\tilde{w}_k}_{\gamma(t_k)} - \tilde{\mathrm{J}}^{\tilde{w}_k}_{\gamma(t_k)}}{h} \right\|_2 \leq D(h^2 + \varepsilon h). \tag{12}$$

Note that this majoration is valid as long as $\tilde{w}_k$ is bounded by a constant which does not depend on k or N, which we have assumed so far.

Gathering (9), (10), (11), and (12), there exists a constant $F > 0$ such that for all k such that $\|w_i\|_2 \leq \|w_0\|_2$ for all $i \leq k$

$$\delta_{k+1} \leq (1+Bh)\delta_k + F(h^2 + h\varepsilon). \tag{13}$$

Combining those inequalities for $k = 1, \ldots, s$, where $s \in \{1, \ldots, N\}$ is such that $\|w_k\|_2 \leq 2\|w_0\|_2$ for all $k \leq s$, we obtain a geometric series whose sum yields

$$\delta_s \leq \frac{F(h^2 + h\varepsilon)}{Bh}(1+Bh)^{s+1}. \tag{14}$$

We now show that for a large enough number of integration steps N, this implies that $\|w_k\|_2 \leq 2\|w_0\|_2$ for all $k \in \{1, \ldots, N\}$. We proceed by contradiction, assuming that there exist arbitrary large $N \in \mathbb{N}$ for which there exists $u(N) \leq N$—that we take to be minimal—such that $\|w_{u(N)}\|_2 > 2\|w_0\|_2$. For any such $N \in \mathbb{N}$, since $u(N)$ is minimal with that property, we can still use (14) with $s = u(N)$:

$$\delta_{u(N)} \leq \frac{F(h^2 + h\varepsilon)}{Bh}(1 + Bh)^{u(N)+1}. \tag{15}$$

Now $h = \frac{1}{N}$ so that

$$\delta_{u(N)} \leq \frac{F(h+\varepsilon)}{B}(1 + Bh)^{u(N)+1} \leq \frac{F(h+\varepsilon)}{B}(1 + Bh)^{\frac{1}{h}+1}. \tag{16}$$

But we have, on the other hand,

$$\|w_0\|_2 < |\|\tilde{w}_{u(N)}\|_2 - \|w_0\|_2| \leq \|\tilde{w}_{u(N)} - w_0\|_2 \leq \frac{F(h+\varepsilon)}{B}(1 + Bh)^{\frac{1}{h}+1}. \tag{17}$$

Taking $\varepsilon \leq h$, which we will keep as an assumption in the rest of the proof, the term on the right goes to zero as $h \to 0$—i.e., as $N \to \infty$— which is a contradiction. So for N large enough, we have that $\|w_k\|_2 \leq 2\|w_0\|_2$ and (14) holds for all $k \in \{1, \ldots, N\}$. With $s = N$, (14) reads as

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

We see that choosing $\varepsilon = \frac{1}{N}$ yields an optimal rate of convergence: choosing a larger value deteriorates the accuracy of the scheme, while choosing a lower value still yields an error in $\mathrm{O}\left(\frac{1}{N}\right)$. Setting $\varepsilon = \frac{1}{N}$,

$$\delta_N \leq \frac{2F}{BN}\left(1 + \frac{B}{N}\right)^{N+1} = \frac{2F}{BN}\left(\exp(B) + \mathrm{o}\left(\frac{1}{N}\right)\right).$$

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

$$\delta_N \leq \frac{G}{N}.$$

□

5. Numerical experiments.

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

- $\mathbb{S}^2$: in spherical coordinates (θ, ϕ), the metric is $g = \left(\begin{smallmatrix} 1 & 0 \\ 0 & \sin(\theta)^2 \end{smallmatrix}\right)$. 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 point of this manifold is the set of symmetric matrices. In [4], the authors endow this space with the following affine-invariant metric: for $\Sigma \in$ SPD(3), $V, W \in$ 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 $\Sigma_0 \in$ SPD(3) with initial tangent vector $X \in$

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

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 [11]. If $\Sigma_0 \in \mathrm{SPD}(3)$ and $X, W \in \mathrm{Sym}(3)$, then

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

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 inverse of the metric. For experiments in large dimensions, we refer the reader to [8].

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 integrate 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. It is, however, a costly operation which requires the computation of the full inverse of the metric at each step.

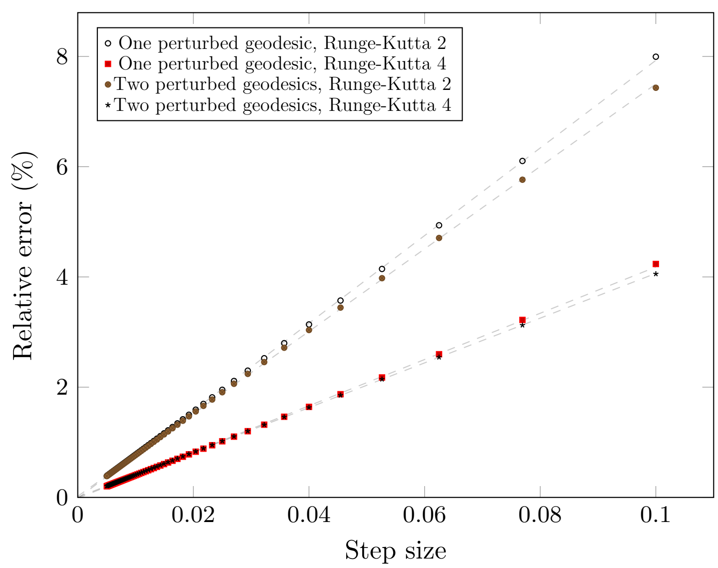

Fig. 3. *Relative errors for the* 2*-sphere in different settings, as functions of the step size, with initial point, velocity, and initial w kept constant. The dotted lines are linear regressions of the measurements. Runge–Kutta* 2 *(resp.,* 4*) indicates that a Runge–Kutta method or order* 2 *(resp.,* 4*) is used for the integration of the geodesic equation.*

5.2. Results. Errors measured in the chosen system of coordinates confirm the linear behavior in both cases, as shown in 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 the 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 using variation 3 of the algorithm, which enforces conservation of the transported vector norm and of its scalar product with the geodesic velocity. Doing so yields an exact transport for the sphere because it is of dimension 2 and the conservation of two quantities is enough to ensure an exact transport—up to the fact that the geodesic is computed approximately—so that the actually observed error is the error in the integration of the geodesic equation. It yields 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 always using 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.

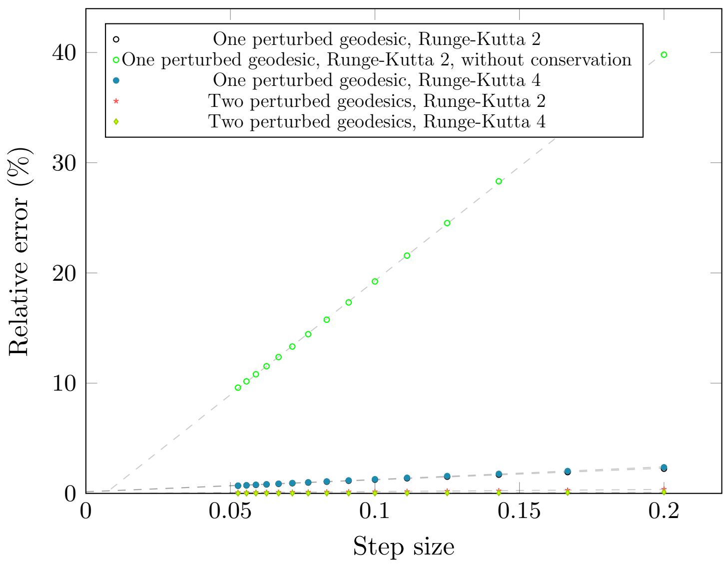

FIG. 4. *Relative errors for SPD(3) in different settings, as functions of the step size, with initial point, velocity, and initial w kept constant. The dotted lines are linear regressions. Runge–Kutta 2 (resp., 4) indicates that a second-order (resp., fourth-order) Runge–Kutta integration has been used to integrate the geodesic equations at steps* (i) *and* (ii)*. Without conservation indicates that variation 3 has not been used.*

We also confirmed numerically that without a second-order method to integrate the geodesic equations at steps (i) and (ii) of the algorithm, the scheme does not converge. This is not in contradiction with Theorem 3.1, which supposes this integration is done with a second-order Runge–Kutta method.

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

5.3. Comparison with Schild's ladder. We compared the relative errors of the fanning scheme with Schild's ladder. We implemented Schild's ladder on the sphere and compared the relative errors of both schemes on the same geodesic and vector. We chose this vector to be orthogonal to the velocity since the transport with Schild's ladder is exact if the transported vector is collinear to the velocity. We

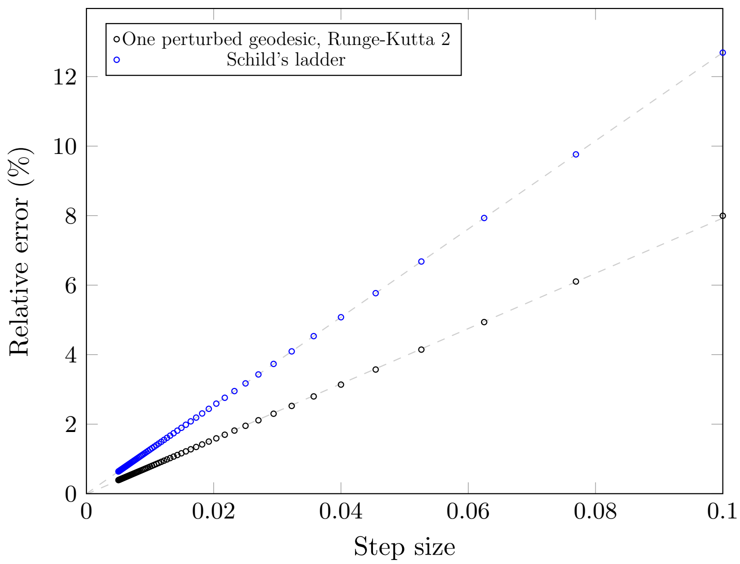

Fig. 5. *Relative error of Schild's ladder scheme compared to the fanning scheme (double geodesic, Runge-Kutta* 2*) proposed here, in the case of* $\mathbb{S}^2$.

use a closed-form expression for the Riemannian logarithm in Schild's ladder and closed-form expressions for the geodesic. The results are given in Figure 5.

6. Conclusion. We proposed a new method, the fanning scheme, to compute parallel transport along a geodesic on a Riemannian manifold using Jacobi fields. In contrast to Schild's ladder, this method does not require the computation of Riemannian logarithms, which may not be given in closed form and may be hard to approximate. We proved that the error of the scheme is of order $\mathrm{O}\left(\frac{1}{N}\right)$, where N is the number of discretization steps, and that it cannot be improved in the general case, yielding the same convergence rate as Schild's ladder. 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.

A limitation of this scheme is to only be applicable when parallel transporting along geodesics, and this limitation seems to be unavoidable with the identity on which it relies. Note also that the Hamiltonian equations are expressed in the cotangent space, whereas the approximation of the transport computed at each step lies in the tangent space to the manifold. Going back and forth from cotangent to tangent space at each iteration is costly if the metric is not available in closed form, as it requires the inversion of a system. In very high dimensions this might limit the performances.

Appendix A. Pseudocode for the algorithm. We give a pseudocode description of the numerical scheme. Here, $G(p)$ denotes the metric matrix at p for any $p \in \mathcal{M}$.

function ParallelTransport(x_0, α_0, w_0, N)
  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

▷ γ_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 $w \in T_{\gamma(0)}\mathcal{M}$
▷ β_0 coordinates of $G(\gamma(0))w_0$
▷ N number of time-steps

$h = 1/N$, $\varepsilon = 1/N$
for $k = 0, \ldots, (N-1)$ **do**
▷ integration of the main geodesic
  $\gamma_{k+\frac{1}{2}} = \gamma_k + \frac{h}{2}v_k$
  $\alpha_{k+\frac{1}{2}} = \alpha_k + \frac{h}{2}\mathrm{F}(\gamma_k, \alpha_k)$
  $\gamma_{k+1} = \gamma_k + h\mathrm{V}(\gamma_{k+\frac{1}{2}}, \alpha_{k+\frac{1}{2}})$
  $\alpha_{k+1} = \alpha_k + h\mathrm{F}(\gamma_{k+\frac{1}{2}}, \alpha_{k+\frac{1}{2}})$
▷ perturbed geodesic equation in the direction w_k
  $\gamma^{\varepsilon}_{k+\frac{1}{2}} = \gamma_k + \frac{h}{2}v(\gamma_k, \alpha_k + \epsilon\beta_k)$
  $\alpha^{\varepsilon}_{k+\frac{1}{2}} = \alpha_k + \varepsilon\beta_k + \frac{h}{2}\mathrm{F}(()\gamma^{\varepsilon}_k, \alpha_k + \varepsilon\beta_k$
  $\gamma^{\varepsilon}_{k+1} = \gamma^{\varepsilon}_k + h\mathrm{V}(\gamma^{\varepsilon}_{k+\frac{1}{2}}, \alpha^{\varepsilon}_k + \frac{1}{2})$
▷ Jacobi field by finite differences
  $\hat{w}_{k+1} = \frac{\gamma^{\varepsilon}_{k+1} - \gamma_{k+1}}{h\varepsilon}$
  $\hat{\beta}_{k+1} = g(\gamma_{k+1})w_{k+1}$ ▷ Use explicit g or solve $K(\gamma_{k+1})\hat{\beta}_{k+1} = \hat{w}_{k+1}$
▷ Conserve quantities
  Solve for a, b:
    $\beta_0^{\top}K(\gamma_0)\beta_0 = (a\hat{\beta}_{k+1} + b\alpha_{k+1})^{\top}K(\tilde{\gamma}_{k+1})(a\hat{\beta}_{k+1} + b\alpha_{k+1}),$
    $\alpha_0^{\top}K(\gamma_0)\alpha_0 = (a\hat{\beta}_{k+1} + b\alpha_{k+1})^{\top}K(\tilde{\gamma}_{k+1})(a\hat{\beta}_{k+1} + b\alpha_{k+1}, v_{k+1})$
  $\beta_{k+1} = a\hat{\beta}_{k+1} + b\alpha_{k+1}$ ▷ parallel transport
  $w_{k+1} = K(\gamma_{k+1})\beta_{k+1}$
end for
return γ_N, α_N, w_N
▷ γ_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

Appendix B. Proofs.

B.1. A lemma to change coordinates. We recall that we suppose the geodesic contained within a compact subset Ω of the manifold $\mathcal{M}$. We start with a result controlling the norms of change-of-coordinates matrices. Let p in $\mathcal{M}$, and let $q = \mathrm{Exp}_p(v)$, where $\|v\|_g \leq \frac{\eta}{2}$, where $\eta > 0$ is a lower bound on the injectivity radius on Ω. We consider two bases of $T_q\mathcal{M}$: one defined from the global system of coordinates, which we denote by B^{Φ}_q, and another made of the normal coordinates centered at p, built from the coordinate on $T_p\mathcal{M}$ obtained from the coordinate chart Φ, which we denote by B^N_q. We can therefore define $\Lambda(p,q)$ as the change-of-coordinates matrix between B^{Φ}_q and B^N_q. The operator norms $|||\cdot|||$ of these matrices are bounded over Ω in the following sense.

LEMMA B.1. *There exists $L \geq 0$ such that for all $p \in K$ and for all $q \in K$ such that $q = \mathrm{Exp}_p(v)$ for some $v \in T_p\mathcal{M}$ with $\|v\|_g \leq \frac{\eta}{2}$, we have*

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

and

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

Proof. Any two norms on $T_q\mathcal{M}$ are equivalent, and the norm bounds of the coordinate change smoothly depend on p and q by smoothness of the metric. Hence we have the result. $\square$

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.

B.2. Transport and connection. We prove a result connecting successive covariant derivatives to parallel transport.

PROPOSITION B.2. *Let V be a vector field on $\mathcal{M}$. Let $\gamma : [0,1] \to \mathcal{M}$ be a geodesic. Then*

$$\nabla^k_{\dot\gamma} V(\gamma(t)) = \left.\frac{\mathrm{d}^k}{\mathrm{d}h^k}\right|_{h=0} P^{-1}_{t,t+h}(V(\gamma(t+h))). \tag{18}$$

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. For $t \in [0,1]$, denote by $(a_i(t))_{i=1,\dots,n}$ the coordinates of $V(\gamma(t))$ in the basis $(E_i(t))_{i=1,\dots,n}$. We have

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

because $P^{-1}_{t,t+h}E_i(t+h) = E_i(t)$ does not depend on h. 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 the definition of $E_i(s)$. $\square$

B.3. A stronger version of Proposition 2.1. From there, we can prove a stronger version of Proposition 2.1. As before, η denotes a lower bound on the injectivity radius of $\mathcal{M}$ on Ω.

PROPOSITION B.3. *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 $h < \frac{\eta}{\|\dot\gamma(t)\|_g}$, we have*

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

Proof. Let $t \in [0,1[$, $w \in T_{\gamma(t)}\mathcal{M}$, and $h < \frac{\eta}{\|\dot\gamma(t)\|_g}$, i.e., such that $J^w_{\gamma(t)}(h)$ is well defined. From Proposition B.2, for any smooth vector field V on $\mathcal{M}$,

$$\nabla^k_{\dot\gamma(t)} V(\gamma(t)) = \left.\frac{\mathrm{d}^k}{\mathrm{d}h^k}\right|_{h=0} P^{-1}_{t,t+h}(V(\gamma(t+h))). \tag{19}$$

We will use this identity to obtain a development of $V(\gamma(t+h)) = J^w_{\gamma(t)}(h)$ for small 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 (1) and finally

$$\begin{aligned}\|\nabla^3_{\dot\gamma} J^w_{\gamma(t)}(h)\|_g &= \|\nabla_{\dot\gamma}(R)(J^w_{\gamma(t)}(h), \dot\gamma(h))\dot\gamma(h) + R(\nabla_{\dot\gamma} J^w_{\gamma(t)}(h), \dot\gamma(h))\dot\gamma(h)\|_g \\ &\leq \|\nabla_{\dot\gamma} R\|_\infty \|\dot\gamma(h)\|_g^2 \|J^w_{\gamma(t)}(h)\|_g + \|R\|_\infty \|\dot\gamma(h)\|_g^2 \|\nabla_{\dot\gamma} J^w_{\gamma(t)}(h)\|_g,\end{aligned} \tag{20}$$

where the ∞-norms, taken over the geodesic and the compact Ω, are finite because the curvature and its derivatives are bounded. Note that we used $\nabla_{\dot\gamma}\dot\gamma = 0$, which holds since γ is a geodesic. In normal coordinates centered at $\gamma(t)$, we have $J^w_{\gamma(t)}(h)^i = hw^i$. Therefore, if we denote by $g_{ij}(\gamma(t+h))$ the components of the metric in normal coordinates, we get using Einstein notations

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

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 Ω. Using Lemma B.1, we get a bound $M \geq 0$ valid on all the systems of normal coordinates centered at a point of the geodesic, so that

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

By the equivalence of the norms as seen in Lemma B.1, and because g varies smoothly, there exists $N \geq 0$ such that

$$\|J^w_{\gamma(t)}(gh)\|_g \leq hMN\|w\|_g, \tag{21}$$

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

$$\|\nabla_{\dot\gamma} J^w_{\gamma(s)}(h)\| \leq C\|w\|_g \tag{22}$$

at any point and for any $h < \max\left(\frac{\eta}{\|\dot\gamma(t)\|_g}, 1-t\right)$. Gathering (20), (21), and (22), we get that there exists a constant $A \geq 0$ which does not depend on t, h, or w such that

$$\left\|\nabla^3_{\dot\gamma} J^w_{\gamma(s)}(h)\right\|_g \leq A\|w\|_g. \tag{23}$$

Now, using (19) with $V(\gamma(t+h)) = J^w_{\gamma(t)}(h)$ and a Taylor formula, we get

$$P^{-1}_{t,t+h}(J^w_{\gamma(t)}(h)) = hw + h^3 r(h, w),$$

where r is the remainder of the expansion, controlled in (23). We thus get

$$\left\|\frac{J^w_{\gamma(t)}(h)}{h} - P_{t,t+h}(w)\right\|_g = \|P_{t,t+h}(h^3 r(w,h))\|_g.$$

Now, because the parallel transport is an isometry, we can use our control (23) on the remainder to get

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

$\square$

B.4. A lemma to control error accumulation. 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 B.4. *There exists $B \geq 0$ such that for all $t \in [0,1[$, for all $w_1, w_2 \in T_{\gamma(t)}\mathcal{M}$, and for all $h \leq \frac{\eta}{\|\dot\gamma(t)\|_g}$ small enough, we have*

$$\left\| \frac{J^{w_1}_{\gamma(t)}(h) - J^{w_2}_{\gamma(t)}(h)}{h} \right\|_2 \leq (1+Bh)\|w_1 - w_2\|_2. \tag{24}$$

Proof. Let $t \in [0,1[$, and let $h \leq \frac{\eta}{\|\dot\gamma(t)\|_g}$. We denote $p = \gamma(t)$, $q = \gamma(t+h)$. We use the exponential map to get normal coordinates on a neighborhood V of p from the basis $\left(\frac{\partial}{\partial x^i}\big|_p\right)_{i=1,\ldots,n}$ of $T_p\mathcal{M}$. Let us denote by $\left(\frac{\partial}{\partial y^i}\big|_r\right)_{i=1,\ldots,n}$ the basis obtained in the tangent space at any point r of V from this system of normal coordinates centered at p. At any point r in V, there are now two different bases of $T_r\mathcal{M}$: $\left(\frac{\partial}{\partial y^i}\big|_r\right)_{i=1,\ldots,n}$ obtained from the normal coordinates and $\left(\frac{\partial}{\partial x^i}\big|_r\right)_{i=1,\ldots,n}$ obtained from the coordinate system Φ. Let $w_1, w_2 \in T_p\mathcal{M}$, and denote by w_j^i for $i \in \{1,\ldots,n\}$, $j \in \{1,2\}$ the coordinates in the global system Φ. By definition, the basis $\left(\frac{\partial}{\partial y^i}\big|_p\right)_{i=1,\ldots,n}$ and the basis $\left(\frac{\partial}{\partial x^i}\big|_p\right)_{i=1,\ldots,n}$ coincide, and in particular, for $j \in \{1,2\}$,

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

If $i \in \{1,\ldots,n\}$, $j \in \{1,2\}$, the jth coordinate of $J^{w_i}_{\gamma(t)}(h)$ in the basis $\left(\frac{\partial}{\partial y^i}\big|_q\right)_{i=1,\ldots,n}$ is

$$J^{w_j}_{\gamma(t)}(h)^i = \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} (\mathrm{Exp}_p(h(v+\varepsilon w_j)))^i = \left.\frac{\partial}{\partial \varepsilon}\right|_{\varepsilon=0} (h(v+\varepsilon w_j))^i = h w_j^i.$$

Let $\Lambda(\gamma(t+h),\gamma(t))$ be the change-of-coordinates matrix of $T_{\gamma(t+h)}$ from the basis $\left(\frac{\partial}{\partial y^i}\big|_q\right)_{i=1,\ldots,n}$ to the basis $\left(\frac{\partial}{\partial x^i}\big|_q\right)_{i=1,\ldots,n}$. Λ varies smoothly with t and h and is the identity when $h=0$. Hence, we can write an expansion

$$\Lambda(\gamma(t+h),\gamma(t)) = \mathrm{Id} + hW(t) + O(h^2).$$

The second-order term depends on the second derivative of Λ with respect to h. Restricting ourselves to a compact subset of $\mathcal{M}$, as in Lemma B.1, we get a uniform bound on the norm of this second derivative, thus getting a control on the operator norm of $\Lambda(\gamma(t+h),\gamma(t))$, which we can write, for h small enough, as

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

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. ∎

B.5. 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 global coordinates, and the error is measured by the 2-norm in these coordinates. $\Phi : \Omega \to U$ denotes the corresponding diffeomorphism. We denote by $\eta > 0$ a lower bound on the injectivity radius of $\mathcal{M}$ on Ω and by $\varepsilon > 0$ the parameter used to compute the perturbed geodesics at step (ii).

PROPOSITION B.5. *There exists $A > 0$ such that for all $t \in [0, 1[$, for all $h \in [0, 1-t]$, and for all $w \in T_{\gamma(t)}\mathcal{M}$,*

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

Proof. Let $t \in [0, 1[$, $h \in [0, 1-t]$, and $w \in T_{\gamma(t)}\mathcal{M}$. The term rewrites as

$$\left\| \frac{\mathrm{J}_{\gamma_k}^{\tilde{w}_k}(h)}{h} - \frac{\mathrm{J}_{\tilde{\gamma}_k}^{\tilde{w}_k}(h)}{h} \right\|_2 = \left\| \frac{\partial \mathrm{Exp}_{\gamma_k}(h\dot{\gamma}_k + x\tilde{w}_k)}{\partial x}\bigg|_{x=0} - \frac{\partial \mathrm{Exp}_{\tilde{\gamma}_k}(h\dot{\tilde{\gamma}}_k + x\tilde{w}_k)}{\partial x}\bigg|_{x=0} \right\|_2 . \tag{25}$$

This is the difference between the derivatives of two solutions of the same differential equation (5) with two different initial conditions. More precisely, we define $\Pi : \Phi(\Omega) \times B_{\mathbb{R}^n}(0, \|\tilde{\gamma}_k\| + 2\varepsilon\|\tilde{w}_k\|) \times [0, \eta]) \to \mathbb{R}^n$ 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 momentum α_0. Π is the flow, in coordinates, of the geodesic equation. We can now rewrite (25) as

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

By the Cauchy–Lipschitz theorem and results on the regularity of the flow, Π is smooth. Hence, its derivatives are bounded over its compact set of definition. Hence there exists a constant $A > 0$ such that

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

where we can once again assume that A is independent of t and h. In coordinates, we use a second-order Runge–Kutta method to integrate the geodesic equation (5) so that the cumulated error $\|\tilde{\gamma} - \gamma\|_2 + \left\|\dot{\tilde{\gamma}} - \dot{\gamma}\right\|_2$ is of order h^2. Hence, there exists a positive constant B which does not depend on h, t, or w such that

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

∎

B.6. Numerical approximation with a single perturbed geodesic. We prove a lemma which allows us to control the error we make when we approximate numerically the Jacobi field using steps (iii) and (ii) of the algorithm.

LEMMA B.6. *For all $L > 0$, there exists $A > 0$ such that for all $t \in [0, 1[$, for all $h \in [0, \frac{\eta}{\|\dot{\gamma}(t)\|_g}]$, and 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 a single perturbed geodesic and a first-order differentiation method.

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

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

$$+ \left\| \underbrace{\frac{\mathrm{Exp}_{\gamma(t)}\big(h(\dot{\gamma}(t) + \varepsilon w)\big) - \mathrm{Exp}_{\gamma(t)}\big(h\dot{\gamma}(t)\big) - \tilde{\mathrm{Exp}}_{\gamma(t)}\big(h(\dot{\gamma}(t) + \varepsilon w)\big) + \tilde{\mathrm{Exp}}_{\gamma(t)}\big(h\dot{\gamma}(t)\big)}{\varepsilon h}}_{(2)} \right\|_2,$$

where Exp is the Riemannian exponential 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.

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

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

This is the error when performing a first-order differentiation on $x \mapsto F^i(x, t, w)$ at 0. This error is of order ϵh and will depend smoothly on t and w. Since $t \in [0, 1]$ and imposing $\|w\|_2 < L$, there exists B which does not depend on t or w such that

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

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

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

Term (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 denote by $x^\varepsilon, \alpha^\varepsilon$ the solution of this equation (in the global system of coordinates) with initial conditions $x^\varepsilon(0) = x_0 = \gamma(t)$ and $\alpha^\varepsilon(0) = \alpha_0^\varepsilon =$

$K(x_0)^{-1}(\dot\gamma(t)+\varepsilon w)$. We denote by $\tilde{x}^\varepsilon$ the result after one step of length h of the integration of the same equation using a second-order Runge–Kutta method with parameter $\delta \in \,]0,1]$. The term (2) rewrites as

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

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

$$x^\varepsilon(h)=x_0+h\dot{x}_0+\frac{h^2}{2}\ddot{x}_0+\int_0^h\frac{(h-t)^2}{2}\dddot{x}^\varepsilon(t)\mathrm{d}t. \tag{26}$$

We have for the last term

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

x^ε being the solution of a smooth ODE 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, and hence there exists $D>0$ such that

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

After computations of the first- and second-order terms, we get

$$\begin{aligned}x^\varepsilon(h)&=x_0+h(\dot\gamma(0)+\varepsilon w)\\&\quad+\frac{h^2}{2}\Big((\nabla_xK)(x_0)[K(x_0)\alpha_0^\varepsilon]\alpha_0^\varepsilon+K(x_0)F_2(x_0,\alpha_0^\varepsilon)\Big)+\mathrm{O}\big(h^3|\varepsilon|\big).\end{aligned} \tag{27}$$

Now we focus on the approximation $\tilde{x}^\varepsilon$. One step of a second-order Runge–Kutta method with parameter δ gives

$$\begin{aligned}\tilde{x}^\varepsilon&=x_0+h\left[\left(1-\frac{1}{2\delta}\right)F_1(x_0,\alpha_0^\varepsilon)+\frac{1}{2\delta}F_1\big(x_0+\delta hF_1(x_0,\alpha_0^\varepsilon),\alpha_0^\varepsilon+\delta hF_2(x_0,\alpha_0^\varepsilon)\big)\right]\\&=x_0+h\left[\left(1-\frac{1}{2\delta}\right)K(x_0)\alpha_0^\varepsilon\ +\frac{1}{2\delta}K\big(x_0+\delta hK(x_0)\alpha_0^\varepsilon\big)\big(\alpha_0^\varepsilon+\delta hF_2(x_0,\alpha_0^\varepsilon)\big)\right].\end{aligned}$$

We use a Taylor expansion for K:

$$\begin{aligned}K\big(x_0+\delta hK(x_0)\alpha_0^\varepsilon\big)&=K(x_0)+\delta h(\nabla_xK)(x_0)[K(x_0)\alpha_0^\varepsilon]\\&\quad+\frac{(\delta h)^2}{2}(\nabla_xK)^2[K(x_0)\alpha_0^\varepsilon,K(x_0)\alpha_0^\varepsilon]+\mathrm{O}\big(h^3\big).\end{aligned}$$

Injecting this into the previous expression for x^ε, we get after development

$$\begin{aligned}\tilde{x}^\varepsilon&=x_0+hK(x_0)(\alpha_0^\varepsilon)\\&\quad+\frac{h^2}{2}\big[K(x_0)F_2(x_0,\alpha_0^\varepsilon)+(\nabla_xK)(x_0)[K(x_0)\alpha_0^\varepsilon]\alpha_0^\varepsilon\big]\\&\quad+\frac{h^3\delta}{4}\big[(\nabla_xK)(x_0)[\alpha_0^\varepsilon]F_2(x_0,\alpha_0^\varepsilon)+(\nabla_xK)^2[K(x_0)\alpha_0^\varepsilon,K(x_0)\alpha_0^\varepsilon]\alpha_0^\varepsilon\big]+\mathrm{O}\big(h^4\big).\end{aligned}$$

The third-order term of $\tilde{x}^\varepsilon - x^0$ is then proportional to

$$(\nabla_x K)(x_0)[\alpha_0^\varepsilon]F_2(x_0, \alpha_0^\varepsilon) - (\nabla_x K)(x_0)\alpha_0^0 F_2(x_0, \alpha_0^0)$$
$$+ (\nabla_x K)^2[K(x_0)\alpha_0^\varepsilon, K(x_0)\alpha_0^\varepsilon]\alpha_0^\varepsilon - (\nabla_x K)^2[K(x_0)\alpha_0^0, K(x_0)\alpha_0^0]\alpha_0^0.$$

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 Ω and tangent vectors in a compact ball in the tangent space, this third-order term is bounded by $Eh^3\varepsilon\|w\|_2$, where E is a positive constant which does not depend on the position on the geodesic. Finally, the zeroth-, first-, and second-order terms of x^ε and $\tilde{x}^\varepsilon$ cancel each other, so that there exists $D \geq 0$ such that

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

which concludes the proof. ∎

B.7. Numerical approximation with two perturbed geodesics. We suppose here that the computation to get the Jacobi field is done using two perturbed geodesics, and a second-order differentiation as described in (8).

LEMMA B.7. *For all $L > 0$, there exists $A > 0$ such that for all $t \in [0, 1[$, for all $h \in [0, 1-t]$, and 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] V. I. Arnol'd, *Mathematical Methods of Classical Mechanics*, Grad. Texts in Math. 60, Springer, 2013.
[2] M. P. do Carmo, *Riemannian Geometry*, Birkhäuser Boston, 1992.
[3] A. Kheyfets, W. A. Miller, and G. A. Newton, *Schild's ladder parallel transport procedure for an arbitrary connection*, Internat. J. Theoret. Phys., 39 (2000), pp. 2891–2898.
[4] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras, *Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing*, J. Math. Imaging Vision, 25 (2006), pp. 423–444.
[5] M. Lorenzi, N. Ayache, and X. Pennec, *Schild's ladder for the parallel transport of deformations in time series of images*, in Proceedings of the Biennial International Conference on Information Processing in Medical Imaging, Springer, 2011, pp. 463–474.
[6] M. Lorenzi and X. Pennec, *Geodesics, parallel transport & one-parameter subgroups for diffeomorphic image registration*, Int. J. Comput. Vis., 105 (2013), pp. 111–127.
[7] M. Lorenzi and X. Pennec, *Parallel transport with pole ladder: Application to deformations of time series of images*, in Geometric Science of Information, Springer, 2013, pp. 68–75.
[8] M. Louis, A. Bône, B. Charlier, S. Durrleman, the A. D. N. Initiative, et al., *Parallel transport in shape analysis: A scalable numerical scheme*, in Geometric Science of Information, Springer, 2017, pp. 29–37.
[9] M. Rumpf and B. Wirth, *Variational time discretization of geodesic calculus*, IMA J. Numer. Anal., 35 (2014), pp. 1011–1046.

[10] J.-B. Schiratti, S. Allassonniere, O. Colliot, and S. Durrleman, *Learning spatiotemporal trajectories from manifold-valued longitudinal data*, in Advances in Neural Information Processing Systems, MIT Press, 2015, pp. 2404–2412.

[11] J.-B. Schiratti, S. Allassonniere, O. Colliot, and S. Durrleman, *A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations*, J. Mach. Learn. Res., 18 (2017), pp. 4840–4872.

[12] L. Younes, *Jacobi fields in groups of diffeomorphisms and applications*, Quart. Appl. Math., 65 (2007), pp. 113–134.

[13] L. Younes, *Shapes and Diffeomorphisms*, Appl. Math. Sci. 171, Springer, 2010.

[14] M. Zhang and P. T. Fletcher, *Probabilistic principal geodesic analysis*, in Advances in Neural Information Processing Systems, MIT Press, 2013, pp. 1178–1186.