



HAL
open science

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

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

► **To cite this version:**

Maxime Louis, Benjamin Charlier, Paul Jusselin, Susovan Pal, Stanley Durrleman. A Fanning Scheme for the Parallel Transport Along Geodesics on Riemannian Manifolds . 2017. hal-01560787v1

HAL Id: hal-01560787

<https://hal.science/hal-01560787v1>

Preprint submitted on 11 Jul 2017 (v1), last revised 7 Jan 2019 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

3 MAXIME LOUIS^{1,2}, BENJAMIN CHARLIER^{3,1,2}, PAUL JUSSELIN^{4,1,2}, SUSOVAN
4 PAL^{1,2}, STANLEY DURRLEMAN^{1,2}

5 ¹INRIA PARIS, ARAMIS PROJECT-TEAM, 75013, PARIS, FRANCE

6 ²SORBONNE UNIVERSITÉS, UPMC UNIV PARIS 06, INSERM, CNRS, INSTITUT DU
7 CERVEAU ET DE LA MOELLE (ICM) - HÔPITAL PITIÉ-SALPÊTRIÈRE, BOULEVARD
8 DE L'HÔPITAL, F-75013, PARIS, FRANCE

9 ³INSTITUT MONTPELLIÉRAIN ALEXANDER GROTHENDIECK, CNRS, UNIV.
10 MONTPELLIER

11 ⁴CMLA, ENS CACHAN

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

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

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

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

44 An alternative is to use the Schild's ladder [1], or its faster version in the case of
45 geodesics the Pole's ladder [5]. These schemes essentially requires the computation
46 of Riemannian exponentials (*Exp*) and logarithms (*Log*) at each step. Usually, the
47 computation of the exponential may be done by integrating Hamiltonian equations,
48 and do not raise specific difficulties. By contrast, the computation of the logarithm
49 must often be done by solving an inverse problem ($Exp \circ Log(x) = x$) with the use of

50 an optimization scheme such as a gradient descent. Such optimization schemes are ap-
 51 proximate and sensitive so the initial conditions and to hyper-parameters, which leads
 52 to additional numerical errors at each step of the scheme. The effects of those numer-
 53 ical errors on the global convergence of the scheme still remain to be studied. When
 54 closed formulas exist for the Riemannian logarithm, or in the case of Lie groups, where
 55 the Logarithm can be approximated efficiently using the Baker-Campbell-Hausdorff
 56 formula (see [4]), the Schild's ladder is an efficient alternative. When this is not the
 57 case, it becomes hardly tractable.

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

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

74 2. Rationale.

75 **2.1. Notations and assumptions.** In this paper, we assume that γ is a geo-
 76 desic defined for all time $t \in [0, 1]$ on a manifold \mathcal{M} of finite dimension $n \in \mathbb{N}$ provided
 77 with the Riemannian metric g . We denote the Riemannian exponential Exp and ∇
 78 the covariant derivative. For $p \in \mathcal{M}$, $T_p\mathcal{M}$ denotes the tangent space of \mathcal{M} at p .
 79 For a vector $w \in T_{\gamma(s)}\mathcal{M}$, for $s, t \in [0, 1]$, we denote $P_{s,t}(w) \in T_{\gamma(t)}\mathcal{M}$ the parallel
 80 transport of w from $\gamma(s)$ to $\gamma(t)$. It is the unique solution at time t of the differential
 81 equation $\nabla_{\dot{\gamma}(u)}P_{s,u}(w) = 0$ for $P_{s,s}(w) = w$. We also note $J_{\gamma(t)}^w(h)$ the Jacobi Field
 82 emerging from $\gamma(t)$ in the direction $w \in T_{\gamma(t)}\mathcal{M}$, that is:

$$83 \quad J_{\gamma(t)}^w(h) = \left. \frac{\partial}{\partial \varepsilon} \right|_{\varepsilon=0} \text{Exp}_{\gamma(t)}(h(\dot{\gamma}(t) + \varepsilon w)) \in T_{\gamma(t+h)}\mathcal{M}$$

84 for $h \in \mathbb{R}$ small enough. It verifies the Jacobi equation (see for instance [6] page
 85 111-119):

$$86 \quad (1) \quad \nabla_{\dot{\gamma}}^2 J_{\gamma(t)}^w(h) + R(J_{\gamma(t)}^w(h), \dot{\gamma}(h))\dot{\gamma}(h) = 0$$

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

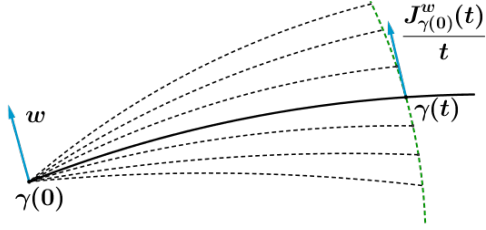


FIGURE 1. The solid line is the geodesic. The green dotted line is formed by the perturbed geodesics at time t . The blue arrows are the initial vector and its approximated parallel transport at time t .

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

97 **2.2. The key identity.** The numerical scheme that we propose arises from the
 98 following identity, which is mentioned in [10]. Figure 1 illustrates the principle.

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

$$100 \quad (2) \quad P_{0,t}(w) = \frac{J_{\gamma(0)}^w(t)}{t} + O(t^2)$$

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

$$105 \quad (3) \quad X^i(t) = w^i + O(t^2).$$

106 By definition, the i -th normal coordinate of $\text{Exp}_{\gamma(0)}(t(v_0 + \varepsilon w))$ is $t(v_0^i + \varepsilon w^i)$. There-
 107 fore, the i -th coordinate of $J_{\gamma(0)}^w(t) = \frac{\partial}{\partial \varepsilon} |_{\varepsilon=0} \text{Exp}_{\gamma(0)}(t(\dot{\gamma}(0) + \varepsilon w))$ is tw^i . Plugging
 108 this into (3) yields the desired result. \square

109 This control on the approximation of the transport by the Jacobi field suggests
 110 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
 111 to approximate the parallel transport of a vector $w \in T_{\gamma(0)}\mathcal{M}$ from $\gamma(0)$ to $\gamma(1)$ by a
 112 sequence of vectors $w_k \in T_{\gamma(\frac{k}{N})}\mathcal{M}$ defined as:

$$113 \quad (4) \quad \begin{cases} w_0 = w \\ w_{k+1} = N J_{\gamma(\frac{k}{N})}^{w_k} \left(\frac{1}{N} \right) \end{cases}$$

114 With the control given in the Proposition 2.1, we can expect to get an error of order
 115 $O(\frac{1}{N^2})$ at each step and hence a speed of convergence in $O(\frac{1}{N})$ overall. There are
 116 manifolds for which the approximation of the parallel transport by Jacobi field is
 117 exact e.g. Euclidean space, but in the general case, one cannot expect to get a better
 118 convergence rate. Indeed, we show in the next Section that this scheme for the sphere
 119 \mathbb{S}^2 has a speed of convergence exactly proportional to $\frac{1}{N}$.

120 **2.3. Convergence rate on \mathbb{S}^2 .** In this Section, we assume that one knows the
 121 geodesic path $\gamma(t)$ and how to compute any Jacobi fields without numerical errors,
 122 and show that the approximation due to Equation (2) alone raises a numerical error
 123 at least of order $O(\frac{1}{N})$.

124 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
 125 the great circles, which may be written as:

$$126 \quad \gamma(t) = \text{Exp}_p(tv) = \cos(t|v|)p + \sin(t|v|)\frac{v}{|v|},$$

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

129 We assume now that $|v| = 1$. Since $w = p \times v$ is orthogonal to v , we have
 130 $\frac{\partial}{\partial \varepsilon} \Big|_{\varepsilon=0} |v + \varepsilon w| = 0$. Therefore:

$$\begin{aligned} 131 \quad \mathbf{J}_p^w(t) &= \frac{\partial}{\partial \varepsilon} \Big|_{\varepsilon=0} \text{Exp}_p(t(v + \varepsilon w)) \\ &= \frac{\partial}{\partial \varepsilon} \Big|_{\varepsilon=0} \left(\cos(t|v + \varepsilon w|)p + \sin(t|v + \varepsilon w|)\frac{v + \varepsilon w}{|v + \varepsilon w|} \right) \\ &= \sin(t)w \end{aligned}$$

132 which does not depend on p . We have $\mathbf{J}_{\gamma(t)}^w(t) = \sin(t)w$. Consequently, the se-
 133 quence of vectors w_k built by the iterative process described in Equation (4) verifies
 134 $w_{k+1} = Nw_k \sin(\frac{1}{N})$ for $k = 0, \dots, N-1$, and $w_N = w_0 N \sin(\frac{1}{N})^N$. In tangent space
 135 coordinates, $P_{0,1}(w_0) = w_0$, so that the numerical error, measured in those tangent
 136 space coordinates, is proportional to $w_0 \left(1 - \left(\frac{\sin(1/N)}{1/N}\right)^N\right)$. We have:

$$137 \quad \left(\frac{\sin(1/N)}{1/N}\right)^N = \exp\left(N \log\left(1 - \frac{1}{6N^2} + o(1/N^2)\right)\right) = 1 - \frac{1}{6N} + o\left(\frac{1}{N}\right)$$

138 yielding:

$$139 \quad \frac{|w_N - w_0|}{|w_0|} \propto \frac{1}{6N} + o\left(\frac{1}{N}\right).$$

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

141 3. The numerical scheme.

142 **3.1. The algorithm.** Unless the metric has some nice properties, there are no
 143 closed forms expressions for the geodesics and the Jacobi fields. Hence, in most
 144 practical cases, these quantities also need to be computed using numerical methods.

145 *Computing geodesics.* In order to avoid the computation of the Christoffel sym-
 146 bols, we propose to integrate the first-order Hamiltonian equations to compute geo-
 147 desics (see [11]). Let $x(t) = (x_1(t), \dots, x_d(t))^T$ be the coordinates of $\gamma(t)$ in a given
 148 local chart, and $\alpha(t) = (\alpha_1(t), \dots, \alpha_d(t))^T$ be the coordinates of the momentum
 149 $g(\gamma(t))\dot{\gamma}(t) \in T_{\gamma(t)}^*\mathcal{M}$ in the same local chart. We have then:

$$150 \quad (5) \quad \begin{cases} \dot{x}(t) = K(x(t))\alpha(t) \\ \dot{\alpha}(t) = -\frac{1}{2}\nabla_x(\alpha(t)^T K(x(t))\alpha(t)) \end{cases},$$

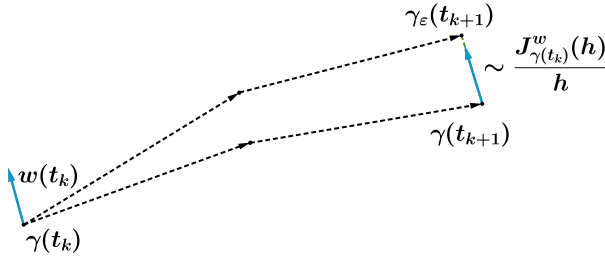


FIGURE 2. One step of the numerical scheme. The dotted arrows represent the steps of the Runge-Kutta integrations for the main geodesic γ and for the perturbed geodesic γ^ϵ . The blue arrows are the initial w and the obtained approximated transport using equation (6).

151 where $K(x(t))$, a d -by- d matrix, is the inverse of the metric g expressed in the local
 152 chart. We will see that to ensure the convergence of the scheme we must use a Runge-
 153 Kutta scheme of order at least 2 to integrate this equation, for which the error is in
 154 $O(\frac{1}{N^2})$.

155 *Computing $J_{\gamma(t)}^w(h)$.* The Jacobi field may be approximated with a numerical
 156 differentiation from the computation of a perturbed geodesic γ^ϵ with initial position
 157 $\gamma(t)$ and initial velocity $\dot{\gamma} + \epsilon w$ where ϵ is a small parameter:

158 (6)
$$J_{\gamma(t)}^w(h) \simeq \frac{\text{Exp}_{\gamma(t)}(h(\dot{\gamma}(t) + \epsilon w)) - \text{Exp}_{\gamma(t)}(h\dot{\gamma}(t))}{\epsilon},$$

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

163 *The algorithm.* Let $N \in \mathbb{N}$. We divide $[0, 1]$ into N intervals $[t_k, t_{k+1}]$, and
 164 initialize with $\gamma_0 = \gamma(0)$, $\dot{\gamma}_0 = \dot{\gamma}(0)$ and $w_0 = w$. The algorithm we propose consists
 165 in iteratively computing, at step k :

- 166 (i) The momentum in the cotangent space corresponding to the vector w_k : $\beta_k =$
 167 $K(\gamma_k)w_k$
- 168 (ii) The new point on the main geodesic γ_{k+1} , by integration of the Hamiltonian
 169 equations using a second-order Runge-Kutta method.
- 170 (iii) The perturbed geodesic starting at γ_k with initial tangent vectors $\dot{\gamma}_k + \epsilon w_k$ at
 171 time h , that we denote γ_{k+1}^ϵ using a second-order Runge-Kutta method.
- (iv) The estimated parallel transport before renormalization :

$$\hat{w}_{k+1} = \frac{\gamma_{k+1}^\epsilon - \gamma_{k+1}}{\epsilon}$$

- (v) The new estimated parallel transport :

$$w_{k+1} = \alpha_k \hat{w}_{k+1} + \beta_k \dot{\gamma}_{k+1}$$

172 where α_k and β_k are normalization factors ensuring $\|w(t_{k+1})\|_g = \|w(t_0)\|_g$ and
 173 $g(w_{k+1}, \dot{\gamma}_{k+1}) = g(w_0, \dot{\gamma}_0)$: those quantities should be conserved during the
 174 transport. This comes at a small cost, and we will see in Proposition 4.2 that
 175 it allows to put a uniform bound on the approximation of the transport by the
 176 Jacobi field.

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

183 **3.2. Order of the approximations and quantity conservations.** As we will
 184 see below, the orders of the different approximations presented above are optimal in
 185 the sense that they are minimal to ensure linear convergence of the scheme. We could
 186 increase the order of the Runge-Kutta integration in the steps (ii) or (iii), or increase
 187 the order of the finite difference approximation of the derivative in step (iii) e.g. by
 188 computing two perturbed geodesics and using a central finite difference:

$$189 \quad J_{\gamma(t)}^w(h) \simeq \frac{\text{Exp}(h(\dot{\gamma}(t) + \varepsilon w)) - \text{Exp}(h(\dot{\gamma}(t) - \varepsilon w))}{2\varepsilon},$$

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

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

200 **3.3. The convergence Theorem.** We obtained the following convergence re-
 201 sult, guaranteeing a linear decrease of the error with the size of the step h .

202 **THEOREM 3.1.** *Let $N \in \mathbb{N}$. Let $w \in T_{\gamma(0)}\mathcal{M}$. We denote $\delta_k = \|P_{0,t_k}(w) - \tilde{w}_k\|_2$
 203 where \tilde{w}_k is the approximate value of the parallel transport of w along γ at time t_k
 204 and where the 2-norm is taken in the coordinates of our global chart. We note ε the
 205 parameter used in the step (iii) and $h = \frac{1}{N}$ the size of the step used of the Runge-Kutta
 206 approximate solution of the geodesic equation.*

207 *With the hypotheses stated in Section 2.1, if we take $\varepsilon = \frac{1}{N}$, then we have:*

$$208 \quad \delta_N = O\left(\frac{1}{N}\right).$$

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

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

220 Before giving a proof of this theorem in Section 5, we prove some lemmas allowing
 221 uniform controls on the different sources of error in the numerical scheme. In Section
 222 4.1, we prove an intermediate results allowing uniform controls on norms of tensors, in
 223 Section 4.2, we prove a stronger result than Proposition 2.1, with stronger hypotheses
 224 and in Section 4.3, we prove a result allowing to control the accumulation of the error.

225 4. Proofs of the lemmas.

226 **4.1. A lemma to change coordinates.** We recall that we suppose the geodesic
 227 contained within a compact subset K of the manifold. We start with a result con-
 228 trolling the norms of change-of-coordinates matrices. Let p in \mathcal{M} and $q \in \mathcal{M}$ within
 229 the radius of the exponential map at p . We consider two basis on $T_p\mathcal{M}$: one defined
 230 from the global system of coordinates, that we note Ψ , and another made of the normal
 231 coordinates (defined from the global system of coordinates Φ) centered at p , that we
 232 note B_p^Ψ . We can therefore define $\Lambda(p, q)$ as the change-of-coordinates matrix between
 233 B_p^Φ and B_p^Ψ . The operators norms $|||\cdot|||$ of these matrices are bounded over K in the
 234 following sense :

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

$$237 \quad |||\Lambda(p, q)||| \leq L$$

238 and

$$239 \quad |||\Lambda^{-1}(p, q)||| \leq L.$$

240 *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
 241 norm is equivalent to the 2-norm $\|\cdot\|_2$ so that there exists $A > 0$ such that for all
 242 $v \in \mathbb{R}^n$, $\|v\|_2 \leq A\|v\|_{g(p)}$. Because K is compact and g varies smoothly, there exists
 243 a constant $A' > 0$ which makes this majoration valid at any point, i.e. such that for
 244 all $p \in \mathcal{M}$, for all $v \in \mathbb{R}^n$, we have :

$$245 \quad (7) \quad \|v\|_2 \leq A'\|v\|_g$$

246 We note $B(0, \frac{\eta}{2A'})$ the closed ball of radius $\frac{\eta}{2A'}$ in $(\mathbb{R}^n, \|\cdot\|_2)$. Let $(p, v) \in K \times B(0, \frac{\eta}{2A'})$.
 247 We note $q = \text{Exp}_p(v)$. The application $\Lambda : (p, v) \rightarrow |||\Lambda(p, v)|||$ is smooth, because
 248 the change of basis matrices smoothly depend on the metric g and on the positions
 249 of p and q . Moreover, Λ is defined on a compact set and hence reaches its maximum
 250 $L \geq 0$. Thanks to the upper bound in (7), when v spans $B(0, \frac{\eta}{2A'})$ in $(\mathbb{R}^n, \|\cdot\|_2)$, it
 251 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
 252 $p \in \mathcal{M}$ and for all q such that $q = \text{Exp}_p(v)$ with $\|v\|_g \leq \frac{\eta}{2}$. We proceed similarly for
 253 Λ^{-1} . \square

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

258 **4.2. A stronger version of Proposition 2.1.** From there, we can prove a
 259 stronger version of Proposition 2.1. We use here the assumption that the manifold
 260 has a strictly positive injectivity radius η on K .

261 PROPOSITION 4.2. *There exists $A \geq 0$ such that for all $t \in [0, 1[$, for all $w \in$
 262 $T_{\gamma(t)}\mathcal{M}$ and for all $h < \max(\frac{\eta}{\|\dot{\gamma}(t)\|_g}, 1 - t)$:*

$$263 \quad \left\| P_{t,t+h}(w) - \frac{J_{\gamma(t)}^w(h)}{h} \right\|_g \leq Ah^2 \|w\|_g.$$

264 *Proof.* Let $t \in [0, 1[$, $w \in T_{\gamma(t)}\mathcal{M}$ and $h < \max(\frac{\eta}{\|\dot{\gamma}(t)\|_g}, 1 - t)$ i.e. such that
 265 $J_{\gamma(t)}^w(h)$ is well defined. The following identity, satisfied for any smooth vector field
 266 V on \mathcal{M} :

$$267 \quad (8) \quad \nabla_{\dot{\gamma}}^k V(\gamma(t)) = \frac{d^k}{dh^k} P_{t,t+h}^{-1}(V(\gamma(t+h)))$$

268 which will be proved in Appendix B.1 provides us with a way to compute the successive
 269 derivatives of $P_{t,t+h}^{-1}(J_{\gamma(t)}^w(h))$.

270 We have $J_{\gamma(t)}^w(0) = 0$, $\nabla_{\dot{\gamma}} J_{\gamma(t)}^w(0) = w$, $\nabla_{\dot{\gamma}}^2 J_{\gamma(t)}^w(0) = -R(J_{\gamma(t)}^w(0), \dot{\gamma}(0))\dot{\gamma}(0) = 0$
 271 using equation (1) and finally:

$$272 \quad (9) \quad \begin{aligned} \|\nabla_{\dot{\gamma}}^3 J_{\gamma(t)}^w(h)\|_g &= \|\nabla_{\dot{\gamma}}(R)(J_{\gamma(t)}^w(h), \dot{\gamma}(h))\dot{\gamma}(h) + R(\nabla_{\dot{\gamma}} J_{\gamma(t)}^w(h), \dot{\gamma}(h))\dot{\gamma}(h)\|_g \\ &\leq \|\nabla_{\dot{\gamma}} R\|_{\infty} \|\dot{\gamma}(h)\|_g^2 \|J_{\gamma(t)}^w(h)\|_g + \|R\|_{\infty} \|\dot{\gamma}(h)\|_g^2 \|\nabla_{\dot{\gamma}} J_{\gamma(t)}^w(h)\|_g, \end{aligned}$$

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

$$277 \quad \|J_{\gamma(t)}^w(h)\|_g^2 = h^2 g_{ij}(\gamma(t+h)) w^i w^j.$$

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

$$282 \quad \|J_{\gamma(t)}^w(g)\|_g \leq hM \|w\|_2.$$

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

$$285 \quad (10) \quad \|J_{\gamma(t)}^w(g)\|_g \leq hMN \|w\|_g$$

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

$$288 \quad (11) \quad \|\nabla_{\dot{\gamma}} J_{\gamma(s)}^w(h)\| \leq C \|w\|_g,$$

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

$$292 \quad (12) \quad \left\| \nabla_{\dot{\gamma}}^3 J_{\gamma(s)}^w(h) \right\|_g \leq A \|w\|_g.$$

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

$$294 \quad J_{\gamma(t)}^w(h) = hP_{t,t+h}(w) + P_{t,t+h}(r(w, h))$$

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

$$296 \quad \left\| \frac{J_{\gamma(t)}^w(h)}{h} - P_{t,t+h}(w) \right\|_g = \|P_{t,t+h}(r(w, h))\|_g.$$

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

$$298 \quad \left\| \frac{J_{\gamma(t)}^w(h)}{h} - P_{t,t+h}(w) \right\|_g \leq \frac{A}{6} h^2 \|w\|_g. \quad \square$$

299 **4.3. A lemma to control the accumulation of the error.** At every step of
300 the scheme, we compute a Jacobi field from an approximate value of the transported
301 vector. We need to control the error made with this computation from an already
302 approximate vector. We provide a control on the 2-norm of the corresponding error,
303 in the global system of coordinates.

304 **LEMMA 4.3.** *There exists $B \geq 0$ such that for all $t \in [0, 1[$, for all $w_1, w_2 \in$
305 $T_{\gamma(t)}\mathcal{M}$, for all $h \leq \frac{\eta}{\|\dot{\gamma}(t)\|_g}$ small enough, we have :*

$$306 \quad (13) \quad \left\| \frac{J_{\gamma(t)}^{w_1}(h) - J_{\gamma(t)}^{w_2}(h)}{h} \right\|_2 \leq (1 + Bh) \|w_1 - w_2\|_2.$$

307 *Proof.* Let $t \in [0, 1[$ and $h \in [0, 1 - t]$. We note $p = \gamma(t)$, $q = \gamma(t + h)$. We use
308 the exponential map to get normal coordinates on a neighborhood of V of p from the
309 basis $\left. \frac{\partial}{\partial x^i} \right|_p$ of $T_p\mathcal{M}$. Let's note $\left. \frac{\partial}{\partial y^i} \right|_r$ the corresponding basis on the tangent space
310 at any point r of V . Let $w_1, w_2 \in T_p\mathcal{M}$ and note w_i^j for $i \in \{1, 2\}$, $j \in \{1, \dots, n\}$
311 the coordinates in the global system. By definition, the basis $\left(\left. \frac{\partial}{\partial y^k} \right|_p \right)$ and the basis
312 $\left(\left. \frac{\partial}{\partial x^k} \right|_p \right)$ coincide, and in particular, for $i \in \{1, 2\}$:

$$313 \quad w_i = (w_i)^k \left. \frac{\partial}{\partial x^k} \right|_p = (w_i)^k \left. \frac{\partial}{\partial y^k} \right|_p.$$

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

$$316 \quad J_{\gamma(t)}^{w_i}(h)^j = \left. \frac{\partial}{\partial \varepsilon} \right|_{\varepsilon=0} (\exp_p(h(v + \varepsilon w_i)))^j = \left. \frac{\partial}{\partial \varepsilon} \right|_{\varepsilon=0} (h(v + \varepsilon w_i))^j = h w^j.$$

317 Let $\Lambda(\gamma(t + h), \gamma(t))$ be the change-of-coordinate matrix of $T_{\gamma(t+h)}$ from the basis
318 $\left(\left. \frac{\partial}{\partial y^k} \right|_q \right)$ to the basis $\left(\left. \frac{\partial}{\partial x^k} \right|_q \right)$. Λ varies smoothly with t and h , and is the identity
319 when $h = 0$. Hence, we can write an expansion :

$$320 \quad \Lambda(\gamma(t + h), \gamma(t)) = Id + hV(t) + O(h^2)$$

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

$$325 \quad \|\Lambda(\gamma(t+h), \gamma(t))\| \leq (1 + Bh)$$

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

$$327 \quad \left\| \frac{J_{\gamma(t)}^{w_1}(h) - J_{\gamma(t)}^{w_2}(h)}{h} \right\|_2 = \|\Lambda(\gamma(t+h), \gamma(t))(w_1 - w_2)\|_2 \leq (1 + Bh) \|w_1 - w_2\|_2$$

328 which is the desired result. \square

329 5. Proof of the convergence Theorem 3.1.

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

$$331 \quad \begin{aligned} \delta_{k+1} &= \|w_{k+1} - \tilde{w}_{k+1}\|_2 \\ &\leq \underbrace{\left\| w_{k+1} - \frac{J_{\gamma_k}^{w_k}(h)}{h} \right\|_2}_{(1)} + \underbrace{\left\| \frac{J_{\gamma_k}^{w_k}(h)}{h} - \frac{J_{\gamma_k}^{\tilde{w}_k}(h)}{h} \right\|_2}_{(2)} \\ &\quad + \underbrace{\left\| \frac{J_{\gamma_k}^{\tilde{w}_k}(h)}{h} - \frac{J_{\tilde{\gamma}_k}^{\tilde{w}_k}(h)}{h} \right\|_2}_{(3)} + \underbrace{\left\| \frac{J_{\tilde{\gamma}_k}^{\tilde{w}_k}(h)}{h} - \frac{\tilde{J}_{\tilde{\gamma}_k}^{\tilde{w}_k}(h)}{h} \right\|_2}_{(4)} \end{aligned}$$

332 where

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

338 We control each of these terms.

339 (1). This is the intrinsic error when using the Jacobi field. We showed in Propo-
 340 sition 4.2 that for h small enough :

$$341 \quad \left\| P_{t_k, t_{k+1}}(w_k) - \frac{J_{\gamma^{(k)}}^{w_k}(h)}{h} \right\|_{g(\gamma(t_{k+1}))} \leq Ah^2 \|w_k\|_g = Ah^2 \|w_0\|_g$$

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

$$344 \quad (14) \quad \left\| P_{t_k, t_{k+1}}(w_k) - \frac{J_{\gamma^{(k)}}^{w_k}(h)}{h} \right\|_2 \leq A'h^2 \|w_0\|_g$$

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

$$346 \quad (15) \quad \left\| \frac{J_{\gamma(t_k)}^{w_k}(h)}{h} - \frac{J_{\gamma(t_k)}^{\tilde{w}_k}(h)}{h} \right\|_2 \leq (1 + Bh)\delta_k$$

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

$$350 \quad (16) \quad \left\| \frac{J_{\gamma_k}^{\tilde{w}_k}(h)}{h} - \frac{\tilde{J}_{\tilde{\gamma}_k}^{\tilde{w}_k}(h)}{h} \right\|_2 \leq Ch^2$$

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

$$357 \quad (17) \quad \left\| \frac{J_{\gamma(t_k)}^{\tilde{w}_k} - \tilde{J}_{\tilde{\gamma}(t_k)}^{\tilde{w}_k}}{h} \right\|_2 \leq D(h^2 + \varepsilon h) \|w_0\|_g.$$

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

360

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

$$363 \quad \delta_{k+1} \leq (1 + Ah)\delta_k + F(h^2 + h\varepsilon).$$

364 Combining those inequalities for $k = 1, \dots, N$, we obtain a geometric series whose sum
 365 yields:

$$366 \quad \delta_N \leq \frac{F(h^2 + h\varepsilon)}{Ah} (1 + Ah)^{N+1}$$

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

$$370 \quad \delta_N \leq \frac{2F}{AN} \left(1 + \frac{A}{N}\right)^{N+1} = \frac{2F}{AN} (\exp(A) + o(\frac{1}{N}))$$

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

$$372 \quad \delta_N \leq \frac{G}{N}. \quad \square$$

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

376 6. Numerical experiments.

377 **6.1. Setup.** We implemented the numerical scheme on simple manifolds where
 378 the parallel transport is known in a closed form, allowing us to evaluate the numerical
 379 error ¹. We present two examples :

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

- 380 • \mathbb{S}^2 : in spherical coordinates (θ, ϕ) the metric is $g = \begin{pmatrix} 1 & 0 \\ 0 & \sin(\theta)^2 \end{pmatrix}$. We gave
 381 expressions for geodesics and parallel transport in Section 2.3.
 382 • The set of 3×3 symmetric positive-definite matrices $\text{SPD}(3)$. The tangent
 383 space at any points of this manifold is the set of symmetric matrices. In [2],
 384 the authors endow this space with the affine-invariant metric: for $\Sigma \in \text{SPD}(3)$,
 385 $V, W \in \text{Sym}(3)$:

$$386 \quad g_{\Sigma}(V, W) = \text{tr}(\Sigma^{-1}V\Sigma^{-1}W)$$

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

$$390 \quad \Sigma(t) = \Sigma_0^{\frac{1}{2}} \exp(tX) \Sigma_0^{\frac{1}{2}}$$

391 where $\exp : \text{Sym}(3) \rightarrow \text{SPD}(3)$ is the matrix exponentiation. Deriving an ex-
 392 pression for the parallel transport can also be done using the explicit Christof-
 393 fel symbols, see [7]. If $\Sigma_0 \in \text{SPD}(3)$ and $X, W \in \text{Sym}(3)$, then :

$$394 \quad P_{0,t}(W) = \exp\left(\frac{t}{2}X\Sigma_0^{-1}\right)W \exp\left(\frac{t}{2}\Sigma_0^{-1}X\right)$$

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

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

403 **6.2. Results.** Errors measured in the chosen system of coordinates confirm the
 404 linear behavior in both cases, as shown on Figures 3 and 4.

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

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

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

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

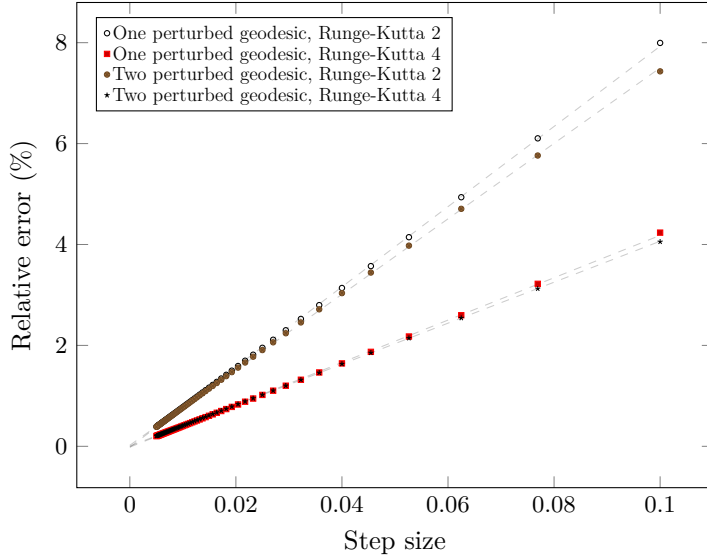


FIGURE 3. Relative error for the 2-Sphere in different settings, as functions of the step size, with initial point, velocity and initial w kept constant. The dotted lines are linear regressions of the measurements.

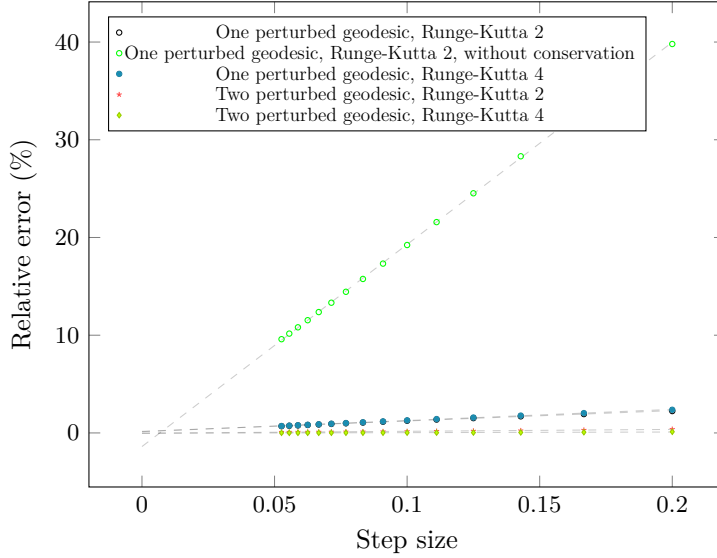


FIGURE 4. Relative errors for SPD(3) in different settings, as functions of the step size, with initial point, velocity and initial w kept constant. The dotted lines are linear regressions.

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

424 **6.3. Comparison with the Schild’s ladder.** We compared the relative errors
 425 of the fanning scheme with the other Christoffel-less method : the Schild’s ladder.
 426 We implemented the Schild’s ladder on the sphere, and compare the relative errors of
 427 both schemes on a same geodesic and vector. We chose this vector to be orthogonal

428 to the velocity, since the transport with the Schild’s ladder is exact if the transported
 429 vector is colinear to the velocity. We use a closed form expression for the Riemannian
 430 logarithm in the Schild’s ladder, and closed form expressions for the geodesic. The
 431 results are given in Figure 5. The fanning scheme is 1.6 times more accurate.

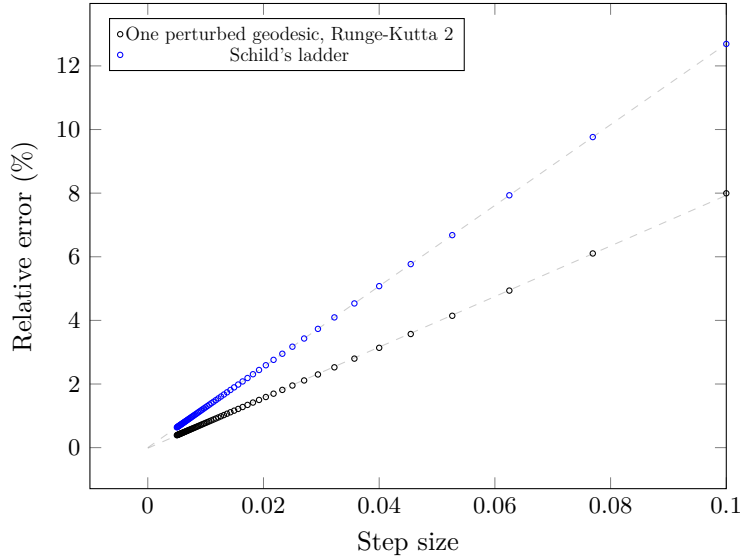


FIGURE 5. Relative error of the Schild’s ladder scheme compared to the fanning scheme (double geodesic, Runge-Kutta 2) proposed here, in the case of \mathbb{S}^2 .

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

433 **7. Conclusion.** We proposed a new method, the fanning scheme, to compute
 434 parallel transport along a geodesic on a Riemannian manifold using Jacobi Fields.
 435 At variance with the Schild’s ladder, this method does not require the computation
 436 of Riemannian logarithms, which are in a lot of cases not given in closed form and
 437 potentially hard to approximate. We proved that the error of the scheme is of order
 438 $O(\frac{1}{N})$ where N is the number of discretization steps, and that it cannot be improved
 439 in the general case, yielding the same convergence rate as the Schild’s ladder. Note
 440 also that, to the best of our knowledge, no convergence result is available for the
 441 Schild’s ladder when extra approximations, which are often necessary, are made –e.g.
 442 approximate Riemannian logarithm through gradient descent or using the Baker-
 443 Haussdorf-Campbell formula. We also showed that only four calls to the Hamiltonian
 444 equations are necessary at each step to provide a satisfying approximation of the
 445 transport, two of them being used to compute the main geodesic. We confirmed the
 446 rate of convergence numerically, and showed empirically that ensuring the conserva-
 447 tions of the norm and of the scalar product with the velocity can yield significant
 448 improvements to the approximation, although this fact still needs to be confirmed in
 449 high dimensions.

450 A limitation of this scheme is to only be applicable when parallel transporting
 451 along geodesics, and an extension to a more general family of curves would be an inter-
 452 esting perspective. Besides, the Hamiltonian equations are expressed in the cotangent
 453 space whereas the velocity lies in the tangent space. Going back and forth from cotan-
 454 gent to tangent space at each iteration can be costly : it typically requires a matrix

455 multiplication, and potentially the inversion of the metric. In very high dimensions
456 this might limit the performances of the scheme.

457 **Acknowledgements.** This work has been partially funded by the European Re-
458 search Council (ERC) under grant agreement No 678304, European Union's Horizon
459 2020 research and innovation programme under grant agreement No 666992, and the
460 program "Investissements d'avenir" ANR-10-IAIHU-06.

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

```

463 1: function PARALLELTRANSPORT( $x_0, \alpha_0, w_0, N$ )
464                                      $\triangleright x_0$  coordinates of  $\gamma(0)$ 
465                                      $\triangleright \alpha_0$  coordinates of  $G(\gamma(0))\dot{\gamma}(0) \in T_{\gamma(0)}^* \mathcal{M}$ 
466                                      $\triangleright w_0$  coordinates of  $w \in T_{\gamma(0)} \mathcal{M}$ 
467                                      $\triangleright N$  number of time-steps
468 2:    $h = 1/N, \varepsilon = 1/N$ 
469 3:   for  $k = 0, \dots, (N - 1)$  do
470                                      $\triangleright$  integration of the main geodesic
471 4:      $x_{k+\frac{1}{2}} = x_k + \frac{h}{2} v_k$ 
472 5:      $\alpha_{k+\frac{1}{2}} = \alpha_k + \frac{h}{2} F(x_k, \alpha_k)$ 
473 6:      $x_{k+1} = x_k + hV(x_{k+\frac{1}{2}}, \alpha_{k+\frac{1}{2}})$ 
474 7:      $\alpha_{k+1} = \alpha_k + hF(x_{k+\frac{1}{2}}, \alpha_{k+\frac{1}{2}})$ 
475                                      $\triangleright$  perturbed geodesic equation in the direction  $w_k$ 
476 8:      $\beta_k = K(x_k)^{-1} w_k$ 
477 9:      $\alpha_k^\varepsilon = \alpha_k + \varepsilon \beta_k$ 
478 10:     $x_{k+\frac{1}{2}}^\varepsilon = x_k + \frac{h}{2} (v_k + \varepsilon w_k)$ 
479 11:     $\alpha_{k+\frac{1}{2}}^\varepsilon = \alpha_k^\varepsilon + \frac{h}{2} F(x_k, \alpha_k^\varepsilon)$ 
480 12:     $x_{k+1}^\varepsilon = x_k^\varepsilon + hV(x_{k+\frac{1}{2}}^\varepsilon, \alpha_{k+\frac{1}{2}}^\varepsilon)$ 
481
482 13:     $J_{k+1} = \frac{x^\varepsilon - x_{k+1}}{\varepsilon}$ 
483                                      $\triangleright$  Jacobi field by finite differences
484                                      $\triangleright$  Conserve quantities
484 14:     $v_{k+1} = V(x_{k+1}, \alpha_{k+1})$ 
485 15:    Solve for  $a, b$  :
486 16:     $G(w_0, w_0) = G(aJ_{k+1} + bv_{k+1}, aJ_{k+1} + bv_{k+1}),$ 
487 17:     $G(v_0, w_0) = G(aJ_{k+1} + bv_{k+1}, v_{k+1})$ 
488 18:     $w_{k+1} = aJ_{k+1} + bv_{k+1}$ 
489                                      $\triangleright$  parallel transport
489 19:  end for
490  return  $x_N, \alpha_N, w_N$ 
491                                      $\triangleright x_N$  approximation of  $\gamma(1)$ 
492                                      $\triangleright \alpha_N$  approximation of  $G(\gamma(1))\dot{\gamma}(1)$ 
493                                      $\triangleright w_N$  approximation of  $P_{\gamma(0), \gamma(1)}(w_0)$ 
494 20: end function
495
496 21: function  $V(x, \alpha)$ 
497 22:   return  $K(x)\alpha$ 
498 23: end function
499
500 24: function  $F(x, \alpha)$ 
501 25:   return  $-\frac{1}{2} \nabla_x (\alpha^T K(x) \alpha)$ 
502                                      $\triangleright$  in closed form or by finite differences
502 26: end function

```


503

504 27: **function** $K(x)$ 505 28: **return** $K(x)$ (or $G(x)^{-1}$)

▷ in closed form

506 29: **end function**507 **Appendix B. Proofs.**508 **B.1. Transport and connection.** We prove a result connecting successive co-
509 variant derivatives to parallel transport:510 **PROPOSITION B.1.** *Let V be a vector field on \mathcal{M} . Let $\gamma : [0, 1] \rightarrow \mathcal{M}$ be a geodesic.*
511 *Then:*

512 (18)
$$\nabla_{\dot{\gamma}}^k V(\gamma(t)) = \frac{d^k}{dh^k} P_{t,t+h}^{-1}(V(\gamma(t+h)))$$

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

515
$$\frac{d^k}{dh^k} P_{t,t+h}^{-1}(V(\gamma(t+h))) = \frac{d^k}{dh^k} P_{t,t+h}^{-1} \sum_{i=1}^n a_i(t+h) E_i(t+h) = \sum_{i=1}^n \frac{d^k a_i(t+h)}{dh^k} E_i(t).$$

516 On the other hand:

517
$$\nabla_{\dot{\gamma}}^k V(\gamma(t)) = \nabla_{\dot{\gamma}}^k \sum_{i=1}^n a_i(t) E_i(t) = \sum_{i=1}^n \nabla_{\dot{\gamma}}^k (a_i(t)) E_i(t) = \sum_{i=1}^n \frac{d^k a_i(t+h)}{dh^k} E_i(t)$$

518 by definition of $E_i(s)$. □519 **B.2. Proof that we can compute the geodesic simultaneously with a**
520 **second-order method.** We give here a control on the error made in the scheme
521 when computing the main geodesic approximately and simultaneously with the par-
522 allel transport. We assume that the main geodesic is computed with a second-order
523 method, and we need to control the subsequent error on the Jacobi field. The com-
524 putations are made in coordinates, and the error measured by the 2-norm on those
525 coordinates.526 **PROPOSITION B.2.** *There exists $A > 0$ such that for all $t \in [0, 1]$, for all $h \in$
527 $[0, 1 - t]$, for all $w \in T_{\gamma(t)}\mathcal{M}$:*

528
$$\left\| \frac{J_{\gamma_k}^{\tilde{w}_k}(h)}{h} - \frac{J_{\tilde{\gamma}_k}^{\tilde{w}_k}(h)}{h} \right\|_2 \leq Ah^2$$

529 *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
530 term rewrites :

531 (19)
$$\left\| \frac{J_{\gamma_k}^{\tilde{w}_k}(h)}{h} - \frac{J_{\tilde{\gamma}_k}^{\tilde{w}_k}(h)}{h} \right\|_2 = \left\| \frac{\partial \text{Exp}_{\gamma}(h\dot{\gamma} + xw)}{\partial x} \Big|_{x=0} - \frac{\partial \text{Exp}_{\tilde{\gamma}}(h\tilde{\dot{\gamma}} + xw)}{\partial x} \Big|_{x=0} \right\|_2$$

532 This is the difference between the derivatives of two solutions of the same differential
533 equation (5) with respect to an initial parameter. More precisely, we define $\Pi :$
534 $\Phi(K) \times B_{\mathbb{R}^n}(0, \|\tilde{\gamma}_k\| + 2\varepsilon \|\tilde{w}_k\|) \times [0, \eta] \rightarrow \mathbb{R}^n$ such that $\Pi(p_0, \alpha_0, h)$ are the coordinates
535 of the solutions of the Hamiltonian equation at time h with initial coordinates p_0 and

536 initial velocity α_0 . Π is the flow, in coordinates, of the geodesic equation. We can
537 now rewrite Equation (19):

$$538 \quad \left\| \frac{\mathbf{J}_{\gamma_k}^{\tilde{w}_k}(h)}{h} - \frac{\mathbf{J}_{\tilde{\gamma}_k}^{\tilde{w}_k}(h)}{h} \right\|_2 = \left\| \frac{\partial \Pi(\gamma_k, \dot{\gamma}_k + \varepsilon \tilde{w}_k, h)}{\partial \varepsilon} \Big|_{\varepsilon=0} - \frac{\partial \Pi(\tilde{\gamma}_k, \dot{\tilde{\gamma}}_k + \varepsilon \tilde{w}_k, h)}{\partial \varepsilon} \Big|_{\varepsilon=0} \right\|_2$$

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

$$542 \quad \left\| \frac{\mathbf{J}_{\gamma_k}^{\tilde{w}_k}(h)}{h} - \frac{\mathbf{J}_{\tilde{\gamma}_k}^{\tilde{w}_k}(h)}{h} \right\|_2 \leq A (\|\tilde{\gamma} - \gamma\|_2 + \|\dot{\tilde{\gamma}} - \dot{\gamma}\|_2)$$

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

$$547 \quad \left\| \frac{\mathbf{J}_{\gamma_k}^{\tilde{w}_k}(h)}{h} - \frac{\mathbf{J}_{\tilde{\gamma}_k}^{\tilde{w}_k}(h)}{h} \right\|_2 \leq Bh^2. \quad \square$$

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

552 **LEMMA B.3.** *For all $L > 0$, There exists $A > 0$ such that for all $t \in [0, 1[$, for all
553 $h \in [0, 1 - t]$, for all $w \in T_{\gamma(t)}\mathcal{M}$ with $\|w\|_2 < L$ -in the global system of coordinates
554 - we have:*

$$555 \quad \left\| \frac{\mathbf{J}_{\gamma(t)}^w(h) - \tilde{\mathbf{J}}_{\gamma(t)}^w(h)}{h} \right\|_2 \leq A(h^2 + \varepsilon h)$$

556 where $\tilde{\mathbf{J}}_{\gamma(t)}^w(h)$ is the numerical approximation of $\mathbf{J}_{\gamma(t)}^w(h)$ computed with a single per-
557 turbed geodesic and a first-order differentiation method. We consider that this approx-
558 imation is computed in the global system of coordinates.

559 *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
560 term in two parts :

$$561 \quad \left\| \frac{\mathbf{J}_{\gamma(t)}^w(h)}{h} - \frac{\tilde{\mathbf{J}}_{\gamma(t)}^w(h)}{h} \right\|_2 \leq \underbrace{\left\| \frac{\mathbf{J}_{\gamma(t)}^w(h)}{h} - \frac{\text{Exp}[h(\dot{\gamma}(t) + \varepsilon w)] - \text{Exp}[h\dot{\gamma}(t)]}{\varepsilon h} \right\|_2}_{(1)} \\ + \underbrace{\left\| \frac{\text{Exp}[h(\dot{\gamma}(t) + \varepsilon w)] - \text{Exp}[h\dot{\gamma}(t)]}{\varepsilon h} - \frac{\tilde{\text{Exp}}[h(\dot{\gamma}(t) + \varepsilon w)] - \tilde{\text{Exp}}[h\dot{\gamma}(t)]}{\varepsilon h} \right\|_2}_{(2)}$$

562 where Exp is the Riemannian exponential at $\gamma(t)$ and $\tilde{\text{Exp}}$ is the numerical approxima-
563 tion of this Riemannian exponential computed thanks to the Hamiltonian equations.
564 When running the scheme, these computations are done in the global system of coor-
565 dinates.

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

$$\begin{aligned}
 & \frac{J_{\dot{\gamma}(t)}^w(h)^i}{h} - \frac{\text{Exp}[h(\dot{\gamma}(t) + \varepsilon w)]^i - \text{Exp}[h\dot{\gamma}(t)]^i}{\varepsilon h} \\
 567 &= \frac{1}{h} \frac{\partial F^i(\varepsilon h, t, w)}{\partial \varepsilon} - \frac{F^i(\varepsilon h, t, w) - F^i(0, w)}{\varepsilon h} \\
 &= \frac{\partial F^i(x, t, w)}{\partial x} \Big|_{x=0} - \frac{F^i(\varepsilon h, t, w) - F^i(0, t, w)}{\varepsilon h}
 \end{aligned}$$

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 $\|w\|_2 \leq L$:

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

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

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

568 (2). We rewrite the Hamiltonian equation $\dot{x}(t) = F_1(x(t), \alpha(t))$ and $\dot{\alpha}(t) =$
 569 $F_2(x(t), \alpha(t))$. We note $x_\varepsilon, \alpha_0^\varepsilon$ the solution of this equation (in the global system
 570 of coordinates) with initial conditions $x_\varepsilon(0) = x_0$ and $\alpha_0^\varepsilon = K(x_0)^{-1}(\dot{\gamma} + \varepsilon w)$. The
 571 term (2) rewrites:

$$572 \quad \frac{1}{\varepsilon h} \|(x^\varepsilon(h) - x^0(h)) - (\tilde{x}^\varepsilon(h) - \tilde{x}^0(h))\|_2$$

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

$$574 \quad (20) \quad x^\varepsilon(h) = x_0 + h\dot{x}^\varepsilon(0) + \frac{h^2}{2}\ddot{x}^\varepsilon(0) + \int_0^h \frac{(h-t)^2}{2} \ddot{\ddot{x}}^\varepsilon(t) dt$$

575 We have, for the last term:

$$576 \quad \left\| \int_0^h \frac{(h-t)^2}{2} \ddot{\ddot{x}}^\varepsilon(t) dt - \int_0^h \frac{(h-t)^2}{2} \ddot{\ddot{x}}^0(t) dt \right\|_2 = \left\| \int_0^h \int_0^{+\varepsilon} \frac{(h-t)^2}{2} \partial_\varepsilon \ddot{\ddot{x}}^\varepsilon(u, t) dt du \right\|_2$$

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

$$580 \quad \left\| \int_0^h \frac{(h-t)^2}{2} \ddot{\ddot{x}}^\varepsilon(t) dt - \int_0^h \frac{(h-t)^2}{2} \ddot{\ddot{x}}^0(t) dt \right\|_2 \leq Dh^3\varepsilon$$

581 For the other terms:

$$582 \quad \dot{x}^\varepsilon(0) = K(x_0)\alpha_0 = \dot{\gamma} + \varepsilon w$$

583 and

$$584 \quad \ddot{x}^\varepsilon(0) = \left. \frac{dK(x^\varepsilon(t))\alpha^\varepsilon(t)}{dt} \right|_{t=0} \\ = (\nabla_x K)(x_0)(\dot{\gamma} + \varepsilon w)\alpha_0^\varepsilon + K(x_0)F_2(x_0, \alpha_0^\varepsilon)$$

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

$$587 \quad \tilde{x}^\varepsilon(h) = x_0 + hF_1\left(x_0 + \frac{h}{2}F_1(x_0, \alpha_0^\varepsilon), \alpha_0^\varepsilon + \frac{h}{2}F_2(x_0, \alpha_0^\varepsilon)\right)$$

588 We replace F_1 and α_0^ε by their expressions:

$$589 \quad \tilde{x}^\varepsilon(h) = x_0 + hK\left(x_0 + \frac{h}{2}F_1(x_0, \alpha_0^\varepsilon)\right)\left(\alpha_0^\varepsilon + \frac{h}{2}F_2(x_0, \alpha_0^\varepsilon)\right) \\ = x_0 + hK\left(x_0 + \frac{h}{2}(\dot{\gamma} + \varepsilon w)\right)\left(\alpha_0^\varepsilon + \frac{h}{2}F_2(x_0, \alpha_0^\varepsilon)\right)$$

590 We use a Taylor expansion for K :

$$591 \quad K\left(x_0 + \frac{h}{2}(\dot{\gamma} + \varepsilon w)\right) = K(x_0) + \frac{h}{2}(\nabla_x K)(x_0)(\dot{\gamma} + \varepsilon w) + \frac{h^2}{8}\nabla^2 K(x_0)(\dot{\gamma} + \varepsilon w) + O(h^3)$$

592 So that:

$$593 \quad \tilde{x}^\varepsilon(h) = x_0 + h(\dot{\gamma} + \varepsilon w) + \frac{h^2}{2}\left[K(x_0)F_2(x_0, \alpha_0^\varepsilon) + \nabla_x K(x_0)(\dot{\gamma} + \varepsilon w)\alpha_0^\varepsilon\right] \\ + \frac{h^3}{4}\left[\nabla_x K(x_0)(\dot{\gamma} + \varepsilon w)F_2(x_0, \alpha_0^\varepsilon) + \frac{1}{2}\nabla^2 K(x_0)(\dot{\gamma} + \varepsilon w)\alpha_0^\varepsilon\right] + O(h^4)$$

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

$$595 \quad \nabla_x K(x_0)\left[(\dot{\gamma} + \varepsilon w)F_2(x_0, \alpha_0^\varepsilon) - (\dot{\gamma})F_2(x_0, \alpha_0^0)\right] \\ + \frac{1}{2}\left[\nabla^2 K(x_0)(\dot{\gamma} + \varepsilon w)\alpha_0^\varepsilon - \nabla^2 K(x_0)(\dot{\gamma})\alpha_0^0\right]$$

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

$$602 \quad \|(x^\varepsilon(h) - x^0(h)) - (\tilde{x}^\varepsilon(h) - \tilde{x}^0(h))\|_2 \leq (Dh^3\varepsilon + Eh^3\varepsilon)\|w\|_g$$

603 which concludes. \square

604 **B.4. Numerical approximation with two perturbed geodesics.** We sup-
605 pose here that the computation to get the Jacobi field is done with a central finite
606 difference method. We prove the following lemma:

607 **LEMMA B.4.** *For all $L > 0$, there exists $A > 0$ such that for all $t \in [0, 1[$, for all
608 $h \in [0, 1 - t]$, for all $w \in T_{\gamma(t)}\mathcal{M}$ with $\|w\|_2 < L$ -in the global system of coordinates
609 - we have:*

$$610 \quad \left\| \frac{J_{\gamma(t)}^w(h) - \tilde{J}_{\gamma(t)}^w(h)}{h} \right\|_2 \leq A(h^2 + \varepsilon h)$$

611 where $\tilde{J}_{\gamma(t)}^w(h)$ is the numerical approximation of $J_{\gamma(t)}^w(h)$ computed with two perturbed
 612 geodesics and a central finite differentiation method. We consider that this approxi-
 613 mation is computed in the global system of coordinates.

614 The proof is similar to the one above.

615

REFERENCES

- 616 [1] Kheyfets, A., Miller, W.A., Newton, G.A.: Schild’s ladder parallel transport procedure for
 617 an arbitrary connection. *International Journal of Theoretical Physics* **39**(12), 2891–2898
 618 (2000)
- 619 [2] Lenglet, C., Rousson, M., Deriche, R., Faugeras, O.: Statistics on the manifold of multivariate
 620 normal distributions: Theory and application to diffusion tensor mri processing. *Journal of*
 621 *Mathematical Imaging and Vision* **25**(3), 423–444 (2006). DOI 10.1007/s10851-006-6897-z.
 622 URL <http://dx.doi.org/10.1007/s10851-006-6897-z>
- 623 [3] Lorenzi, M., Ayache, N., Pennec, X.: Schild’s ladder for the parallel transport of deformations
 624 in time series of images. In: *Biennial International Conference on Information Processing*
 625 *in Medical Imaging*, pp. 463–474. Springer (2011)
- 626 [4] Lorenzi, M., Pennec, X.: Geodesics, Parallel Transport & One-parameter Subgroups for Diffeo-
 627 morphic Image Registration. *International Journal of Computer Vision* **105**(2), 111–127
 628 (2013). DOI 10.1007/s11263-012-0598-4. URL <https://hal.inria.fr/hal-00813835>
- 629 [5] Lorenzi, M., Pennec, X.: Parallel Transport with Pole Ladder: Application to Deforma-
 630 tions of time Series of Images. In: F. Nielsen, F. Barbaresco (eds.) *GSI2013 - Geomet-*
 631 *ric Science of Information*, vol. 8085, pp. 68–75. Springer, Paris, France (2013). DOI
 632 10.1007/978-3-642-40020-9_6. URL <https://hal.inria.fr/hal-00819898>
- 633 [6] Manfredo, P.: *Riemannian geometry* (1992)
- 634 [7] Schiratti, J.B.: *Methods and algorithms to learn spatio-temporal changes from longitudinal*
 635 *manifold-valued observations*. PhD Thesis (2017)
- 636 [8] Schiratti, J.B., Allasonnière, S., Colliot, O., Durrleman, S.: Learning spatiotemporal trajec-
 637 tories from manifold-valued longitudinal data. In: *NIPS 28*
- 638 [9] Vogtmann, K., Weinstein, A., Arnol’d, V.: *Mathematical Methods of Classical Mechanics*.
 639 *Graduate Texts in Mathematics*. Springer New York (1997). URL [https://books.google.](https://books.google.fr/books?id=Pd8-s6rOt_cC)
 640 [fr/books?id=Pd8-s6rOt_cC](https://books.google.fr/books?id=Pd8-s6rOt_cC)
- 641 [10] Younes, L.: Jacobi fields in groups of diffeomorphisms and applications. *Quarterly of Applied*
 642 *Mathematics* **65**(1), 113–134 (2007). URL <http://www.jstor.org/stable/43638762>
- 643 [11] Younes, L.: *Shapes and diffeomorphisms*. Heidelberg : Springer (2010). ”A direct application
 644 of what is presented in the book is a branch of the computerized analysis of medical images
 645 called computational anatomy”–Back cover
- 646 [12] Zhang, M., Fletcher, P.: Probabilistic principal geodesic analysis. In: C.J.C. Burges, L. Bot-
 647 tou, M. Welling, Z. Ghahramani, K.Q. Weinberger (eds.) *Advances in Neural Inform-*
 648 *ation Processing Systems 26*, pp. 1178–1186. Curran Associates, Inc. (2013). URL
 649 <http://papers.nips.cc/paper/5133-probabilistic-principal-geodesic-analysis.pdf>