On Riemannian Stochastic Approximation Schemes with Fixed Step-Size
Supplementary Material

Alain Durmus
Université Paris-Saclay
ENS Paris-Saclay, CNRS
Centre Borelli, F-91190 Gif-sur-Yvette, France
alain.durmus@ens-paris-saclay.fr

Pablo Jiménez
CMAP, École Polytechnique
Institut Polytechnique de Paris
pablo.jimenez-moreno@polytechnique.edu

Éric Moulines
CMAP, École Polytechnique, CNRS
Institut Polytechnique de Paris
eric.moulines@polytechnique.edu

Salem Said
Laboratoire IMS, Université de Bordeaux
salem.said@u-bordeaux.fr

Contents

S1 Assumptions

On the manifold

A1. *Assume one of the following conditions.*

(i) Θ is a Hadamard manifold, i.e. a complete, simply connected Riemannian manifold with non-positive sectional curvature. In addition, S is a closed geodesically convex subset of Θ with non-empty interior.
(ii) Θ is a complete, connected Riemannian manifold and $\mathsf{S} = \Theta$.

A2. *Θ is a Hadamard manifold. In addition, there exists $\kappa > 0$ such that the sectional curvature of Θ is bounded below by $-\kappa^2$.*

On the distribution of the data

MD1. *The sequence $(X_n)_{n\in\mathbb{N}^*}$ is independent and identically distributed (i.i.d.). In addition, for any $\theta \in \Theta$, $\mathbb{E}[e_\theta(X_1)] = 0$ and there exist $\sigma_0^2, \sigma_1^2 > 0$ such that for any $\theta \in \mathsf{S}$, $\mathbb{E}[\|e_\theta(X_1)\|_\theta^2] \leq \sigma_0^2 + \sigma_1^2 \|h(\theta)\|_\theta^2$.*

MD2. *(i) $\mathbb{P}$-almost surely, the vector field $\theta \mapsto e_\theta(X_1)$ is continuous on Θ.*
(ii) For any $\theta \in \Theta$, Leb_θ and the distribution of $e_\theta(X_1)$ are mutually absolutely continuous.

MD3. *Σ is a continuous tensor field of type $(2,0)$ on Θ.*

MD4. *There exist $\varepsilon_e > 0$, $\tilde{\sigma}_0^2, \tilde{\sigma}_1^2 \geq 0$ such that for any $\theta \in \Theta$, $\mathbb{E}[\|e_\theta(X_1)\|_\theta^{2+\varepsilon_e}] \leq \tilde{\sigma}_0^2 + \tilde{\sigma}_1^2 V(\theta)$.*

MD5. *There exists $\theta \in \Theta$ such that*

$$\int_\Theta \rho_\Theta^2(\theta, \nu)\pi(\mathrm{d}\nu) < +\infty .$$

On the Lyapunov function V and the mean field function h

H1. *(i) For any $\theta \in \Theta$, $V \circ \mathrm{proj}_{\mathsf{S}}(\theta) \leq V(\theta)$.*
(ii) V is continuously differentiable on Θ and its Riemannian gradient $\mathrm{grad}\, V$ is geodesically L-Lipschitz, i.e., there exists $L \geq 0$ such that for any $\theta_0, \theta_1 \in \Theta$, and geodesic curve $\gamma : [0,1] \to \Theta$ such that $\gamma(0) = \theta_0$ and $\gamma(1) = \theta_1$,

$$\|\mathrm{grad}\, V(\theta_1) - \mathrm{T}_{01}^\gamma \mathrm{grad}\, V(\theta_0)\|_{\theta_1} \leq L\ell(\gamma) ,$$

where $\ell(\gamma) = \|\dot\gamma(0)\|_{\theta_0}$ is the length of the geodesic.
(iii) V is proper on S, i.e., for any $M \geq 0$, there exists a compact set $\mathsf{K} \subset \mathsf{S}$ such that for any $\theta \in \mathsf{S} \setminus \mathsf{K}$, $V(\theta) > M$.

H2. *There exist $C_1 \geq 0$ and $C_2 > 0$ such that for any $\theta \in \mathsf{S}$, $\|h(\theta)\|_\theta^2 + C_2 \langle \mathrm{grad}\, V(\theta), h(\theta)\rangle_\theta \leq C_1$.*

H3 $(\mathsf{K}^\star)$**.** *There exists $\lambda > 0$ such that for any $\theta \in \mathsf{S}$, $\langle \mathrm{grad}\, V(\theta), h(\theta)\rangle_\theta \leq -\lambda V(\theta)\mathbb{1}_{\mathsf{S}\setminus\mathsf{K}^\star}(\theta)$.*

H4. *There exists $\theta^\star \in \mathsf{S}$ such that for any $r > 0$, **H**3$(\overline{\mathrm{B}}(\theta^\star, r))$ holds and that there exists $c_r > 0$ satisfying for any $\theta \in \mathsf{S} \setminus \overline{\mathrm{B}}(\theta^\star, r)$, $c_r \leq V(\theta)$.*

H5. *There exist a linear mapping $\mathbf{A} : \mathrm{T}_{\theta^*}\Theta \to \mathrm{T}_{\theta^*}\Theta$ and a map $\mathscr{H} : \Theta \to \mathrm{T}_{\theta^*}\Theta$, such that for any $\theta \in \Theta$,*

$$h(\theta) = \mathrm{T}_{01}^\gamma \left(\mathbf{A}\mathrm{Exp}_{\theta^\star}^{-1}(\theta) + \mathscr{H}(\theta)\right) ,$$

*where $\theta^\star$ is defined in **H**4, T_{01}^γ denotes parallel transport along the geodesic $\gamma : [0,1] \to \Theta$ with $\gamma(0) = \theta^\star$ and $\gamma(1) = \theta$, and $\lim_{\theta\to\theta^\star}\{\|\mathscr{H}(\theta)\|_{\theta^\star} / \rho_\Theta(\theta^\star, \theta)\} = 0$. In addition, the eigenvalues of the matrix $\mathbf{A}$ all have strictly negative real parts. Finally, there exists $C_3 > 0$ such that for any $\theta \in \Theta$, $\|h(\theta)\|_\theta \leq C_3 \rho_\Theta(\theta^\star, \theta)$.*

H6. *There exists $\theta^\star$ such that **H**3$(\{\theta^\star\})$ holds and there exists $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ such that for any $\theta \in \Theta$, $V(\theta) \geq \phi(\rho_\Theta(\theta^\star, \theta))$ and for any $r > 0$, $\inf_{[r,+\infty)} \phi > 0$. In addition, there exists $\bar{a} > 0$, such that $\lim_{r\to+\infty} \sup_{a\leq\bar{a}} a/\phi(a^{1/2}r) = 0$.*

On the objective function f in the gradient case

F1. *$f : \Theta \to \mathbb{R}$ is twice continuously differentiable and $\operatorname{grad} f$ is geodesically L_f-Lipschitz, see (3).*

F 2. *f is λ_f-strongly geodesically convex, for some $\lambda_f > 0$, i.e. for any $\theta_1, \theta_2 \in \Theta$, $f(\theta_2) \geq f(\theta_1) + \langle \operatorname{Exp}_{\theta_1}^{-1}(\theta_2), \operatorname{grad} f(\theta_1)\rangle_{\theta_1} + \lambda_f \rho_\Theta^2(\theta_1, \theta_2)/2$.*

F 3. *f is twice continuously differentiable. There exists $\tilde{\lambda}_f > 0$ such that for any $\theta \in \Theta$, $-\langle \operatorname{Exp}_\theta^{-1}(\theta^\star), \operatorname{grad} f(\theta)\rangle_\theta \geq \tilde{\lambda}_f V_1(\theta)$, where V_1 is defined by (9) with $\delta = 1$. In addition, there exists $C_f > 0$ such that for any $\theta \in \Theta$, $\|\operatorname{grad} f(\theta)\|_\theta^2 \leq C_f(\rho_\Theta^2(\theta^\star, \theta) \wedge 1)$.*

S2 Supplementary notation

Denote the unit tangent space $\mathrm{U}_\theta\Theta = \{u \in \mathrm{T}_\theta\Theta : \|u\|_\theta = 1\}$. The cut-locus of θ, $\operatorname{Cut}(\theta) \subset \Theta$ [1, p. 308] and the injectivity domain $\operatorname{ID}(\theta) \subset \mathrm{T}_\theta\Theta$ [1, p. 310] are two notions that inform us about the length-minimizing properties of geodesics, and therefore provide the domain of definition of the Riemannian exponential. On a complete and connected manifold, [1, Theorem 10.34] holds, meaning the restriction $(\operatorname{Exp}_\theta)_{|\operatorname{ID}(\theta)} : \operatorname{ID}(\theta) \to \Theta$ is a diffeomorphism onto its image $\Theta \setminus \operatorname{Cut}(\theta)$. We simply denote $\operatorname{Exp}_\theta^{-1} : \Theta \setminus \operatorname{Cut}(\theta) \to \operatorname{ID}(\theta)$ its inverse. Under the assumption that Θ is complete, simply connected and of non-positive sectional curvature, *i.e.* a Hadamard manifold, [1, Proposition 12.9] proves that $\operatorname{Cut}(\theta) = \emptyset$ and $\operatorname{ID}(\theta) = \mathrm{T}_\theta\Theta$ for any $\theta \in \Theta$.

For a measure μ on a measurable space $(\mathsf{Y}, \mathcal{Y})$, denote by $\mu(g)$ the integral of a measurable function $g : \mathsf{Y} \to \mathbb{R}$ with respect to μ, when it exists.

S3 Proofs of Section 2

Under **A**1 and **MD**1, for any $\eta > 0$, we denote by Q_η the Markov kernel associated with $(\theta_n)_{n\in\mathbb{N}}$ defined by (2) given for any $\mathsf{A} \in \mathcal{B}(\mathsf{S})$ and $\theta \in \mathsf{S}$ by

$$Q_\eta(\theta, \mathsf{A}) = \mathbb{E}\left[\mathbb{1}_{\mathsf{A}}\left(\operatorname{Exp}_\theta\{\eta H_\theta(X_1)\}\right)\right] . \tag{S1}$$

Useful notions, definitions and results relative to Markov chain theory are given in Section S6.1.

Lemma S1. *Assume **A**1, **MD**1, **H**1-(i)-(ii). Then for any $\eta > 0$ and $\theta_0 \in \mathsf{S}$,*

$$Q_\eta V(\theta_0) \leq V(\theta_0) + \eta \langle \operatorname{grad} V(\theta_0), h(\theta_0)\rangle_{\theta_0} + L\eta^2 \left[\|h(\theta_0)\|_{\theta_0}^2 + \sigma_0^2 + \sigma_1^2 \|h(\theta_0)\|_{\theta_0}^2\right] . \tag{S2}$$

Proof. Let $\theta_0 \in \mathsf{S}$, and $\eta > 0$. Consider

$$\theta_{1/2} = \operatorname{Exp}_{\theta_0}\left[\eta H_{\theta_0}(X_1)\right] , \; \theta_1 = \operatorname{proj}_{\mathsf{S}}\left(\theta_{1/2}\right) . \tag{S3}$$

First, by definition of Q_η and **H**1-(i), we have

$$Q_\eta V(\theta_0) = \mathbb{E}\left[V(\theta_1)\right] \leq \mathbb{E}\left[V(\theta_{1/2})\right] . \tag{S4}$$

Second, using **A**1, **H**1-(ii), [2, Lemma 1] and (S3), we obtain

$$V(\theta_{1/2}) \leq V(\theta_0) + \eta \langle \operatorname{grad} V(\theta_0), H_{\theta_0}(X_1)\rangle_{\theta_0} + (L/2)\eta^2 \|H_{\theta_0}(X_1)\|_{\theta_0}^2 .$$

Plugging this result in (S4) and using **MD**1 completes the proof of (S2). □

S3.1 Proof of Theorem 1

(a) Using Lemma S1 and **H**2 we have for any $\theta_0 \in \mathsf{S}$ and $\eta > 0$,

$$Q_\eta V(\theta_0) \leq V(\theta_0) + \eta\{1 - C_2 L\eta(1+\sigma_1^2)\} \langle \operatorname{grad} V(\theta_0), h(\theta_0)\rangle_{\theta_0} + L\eta^2[\sigma_0^2 + C_1(1+\sigma_1^2)] .$$

Letting $\overline{\eta} = [2C_2 L(1+\sigma_1^2)]^{-1}$, then for any $\eta \in (0, \overline{\eta}]$, we have $1 - C_2 L\eta(1+\sigma_1^2) \geq 1/2$. Therefore, using also that $\langle \operatorname{grad} V(\theta_0), h(\theta_0)\rangle_{\theta_0} \leq 0$, we obtain,

$$Q_\eta V(\theta_0) \leq V(\theta_0) + (\eta/2) \langle \operatorname{grad} V(\theta_0), h(\theta_0)\rangle_{\theta_0} + L\eta^2[\sigma_0^2 + C_1(1+\sigma_1^2)] . \tag{S5}$$

Therefore, by the Markov property, for any $k \in \mathbb{N}^*$, $\eta \in (0, \bar{\eta}]$ and $\theta_0 \in \mathsf{S}$ we get,

$$-(\eta/2) \int_\Theta \langle \operatorname{grad} V(\theta), h(\theta) \rangle_\theta \, Q_\eta^{k-1}(\theta_0, \mathrm{d}\theta) \le Q_\eta^{k-1} V(\theta_0) - Q_\eta^k V(\theta_0) + L\eta^2[\sigma_0^2 + C_1(1+\sigma_1^2)] \, .$$

Summing these inequalities for $k \in \{1, \dots, n\}$ concludes the proof of (a) upon using that V is a non-negative function.
(b) We prove (5) by using **H**3($\mathsf{K}^\star$) in (4) and dividing both sides by $\lambda > 0$.
(c) We start by using **H**3($\mathsf{K}^\star$) in (S5). For any $\eta \in (0, \bar{\eta}]$ and $\theta_0 \in \mathsf{S}$, we have

$$Q_\eta V(\theta_0) \le V(\theta_0) \left[1 - (\lambda\eta/2) \mathbb{1}_{\mathsf{S} \setminus \mathsf{K}^\star}(\theta_0)\right] + \eta^2 b/2 \, , \tag{S6}$$

where $b = 2L[\sigma_0^2 + C_1(1+\sigma_1^2)]$. By adding and subtracting $V(\theta_0)(\lambda\eta/2)\mathbb{1}_{\mathsf{K}^\star}(\theta_0)$ in the right-hand side of (S6), we have,

$$Q_\eta V(\theta_0) \le V(\theta_0)[1 - \eta a] + \eta(b\eta/2 + a \|V\|_{\mathsf{K}^\star}) \, , \tag{S7}$$

where $a = \lambda/2$. Therefore, by a straightforward induction on $n \in \mathbb{N}$, using the Markov property, we get, for any $n \in \mathbb{N}$, $\eta \in (0, \bar{\eta}]$ and $\theta_0 \in \mathsf{S}$,

$$\begin{aligned}\mathbb{E}[V(\theta_n)] &\le \{1 - \eta a\}^n V(\theta_0) + \eta(b\eta/2 + a\|V\|_{\mathsf{K}^\star}) \sum_{k=0}^{n-1} [1 - \eta a]^k \\ &\le \{1 - \eta a\}^n V(\theta_0) + \{\|V\|_{\mathsf{K}^\star} + (b\eta/2a)\} \, ,\end{aligned}$$

which concludes the proof of (c) and Theorem 1.

S3.2 An alternative to Theorem 1-(b)

Consider the following condition for some compact set $\mathsf{K}^\star \subset \mathsf{S}$.

HS1 ($\mathsf{K}^\star$). *There exists $\lambda > 0$ such that for any $\theta \in \mathsf{S}$, $\langle \operatorname{grad} V(\theta), h(\theta) \rangle_\theta \le -\lambda \|h(\theta)\|_\theta^2 \mathbb{1}_{\mathsf{S} \setminus \mathsf{K}^\star}(\theta)$.*

Theorem S2. *Assume **A**1, **MD**1, **H**1-(i)-(ii) and **HS**1($\mathsf{K}^\star$) hold for some compact set $\mathsf{K}^\star \subset \mathsf{S}$, and define $\|h\|_{\mathsf{K}^\star} = \sup\{\|h(\theta)\|_\theta : \theta \in \mathsf{K}^\star\}$ if $\mathsf{K}^\star \neq \emptyset$ and $\|h\|_{\mathsf{K}^\star} = 0$ otherwise. Then for any $\eta \in (0, \check{\eta}]$ and $\theta_0 \in \mathsf{S}$, and $n \in \mathbb{N}^*$,*

$$n^{-1} \sum_{k=0}^{n-1} \mathbb{E}[\mathbb{1}_{\mathsf{S} \setminus \mathsf{K}^\star}(\theta_k) \|h(\theta_k)\|_{\theta_k}^2] \le V(\theta_0)/(an\eta) + \eta \tilde{b}/a \, ,$$

where $(\theta_n)_{n \in \mathbb{N}}$ is defined by (2) starting from θ_0, $\check{\eta} = \lambda/[2(1+\sigma_1^2)L]$, $a = \lambda/2$ and $\tilde{b} = L((1+\sigma_1^2)\|h\|_{\mathsf{K}^\star} + \sigma_0^2)$.

Proof. By Lemma S1 and **HS**1($\mathsf{K}^\star$), for any $\eta \in (0, \bar{\eta}]$ and $\theta_0 \in \mathsf{S}$, we have

$$Q_\eta V(\theta_0) \le V(\theta_0) - \eta\lambda \|h(\theta_0)\|_{\theta_0}^2 \mathbb{1}_{\mathsf{S} \setminus \mathsf{K}^\star}(\theta_0) + L\eta^2 \left[\|h(\theta_0)\|_{\theta_0}^2 + \sigma_0^2 + \sigma_1^2 \|h(\theta_0)\|_{\theta_0}^2\right] \, .$$

Therefore, by the Markov property, for any $k \in \mathbb{N}^*$, $\eta \in (0, \bar{\eta}]$ and $\theta_0 \in \mathsf{S}$, we get

$$\begin{aligned}&(\eta\lambda/2) \int_\Theta \{\mathbb{1}_{\mathsf{S} \setminus \mathsf{K}^\star}(\theta) \|h(\theta)\|_\theta^2\} Q_\eta^{k-1}(\theta_0, \mathrm{d}\theta) \\ &\qquad \le Q_\eta^{k-1} V(\theta_0) - Q_\eta^k V(\theta_0) + L\eta^2((1+\sigma_1^2)\|h\|_{\mathsf{K}^\star} + \sigma_0^2) \, .\end{aligned}$$

Summing these inequalities for $k \in \{1, \dots, n\}$ concludes the proof upon using that V is a non-negative function. □

S3.3 Proof of Theorem 2

Lemma S3. *Assume **A**1, **MD**1 and **MD**2-(i). Then the Markov kernel Q_η on $\mathsf{S} \times \mathcal{B}(\mathsf{S})$ is Feller, i.e. for any measurable bounded function $f : \mathsf{S} \to \mathbb{R}$, $Q_\eta f$ is continuous from S to $\mathbb{R}$.*

Proof. The proof is an easy consequence of the Lebesgue dominated convergence theorem, since h is continuous and **MD**2-(i) holds. □

For the next lemma, we introduce μ_{S}, the restriction to S of the Riemannian measure μ_Θ associated with the volume form on Θ.

Lemma S4. *Assume **A**1, **MD**1 and **MD**2-(ii). Then Q_η is μ_{S}-irreducible and aperiodic.*

Proof. We consider first the case **A**1-(i), where Θ is a Hadamard manifold. Let $\mathsf{A} \in \mathcal{B}(\mathsf{S})$ be a Borel set of S, such that $\mu_{\mathsf{S}}(\mathsf{A}) > 0$. We only need to show that for any $\theta_0 \in \Theta$, $Q_\eta(\theta_0, \mathsf{A}) > 0$. Indeed, this gives μ_{S}-irreducibility by definition and implies that the chain is aperiodic by [3, Theorem 5.4.4] since for any $\mathsf{A} \in \mathcal{B}(\mathsf{S})$, $\mu_{\mathsf{S}}(\mathsf{A}) > 0$, $\theta \in \mathsf{A}$, we have $Q_\eta(\theta, \mathsf{A}) > 0$.

Let $\theta_0 \in \mathsf{S}$. By definition of the scheme (2) and $\mathrm{proj}_{\mathsf{S}}$, $Q_\eta(\theta_0, \mathsf{A}) = \mathbb{P}(\mathrm{proj}_{\mathsf{S}} \circ \mathrm{Exp}_{\theta_0}(\eta\{h(\theta_0) + e_{\theta_0}(X_1)\}) \in \mathsf{A}) \geq \mathbb{P}(\mathrm{Exp}_{\theta_0}(\eta\{h(\theta_0) + e_{\theta_0}(X_1)\}) \in \mathsf{A})$. However, using **MD**2-(ii), the law of $e_{\theta_0}(X_1)$ has a positive density $\phi : \mathrm{T}_{\theta_0}\Theta \to (0, +\infty)$ with respect to Lebesgue's measure Leb_{θ_0}. Denote $(\mathfrak{g}_{ij}(\theta))_{1 \leq i,j \leq d}$ the matrix representing the Riemannian metric at $\theta \in \Theta$ in normal global coordinates at θ_0. Expressing μ_{S} in these coordinates and using [1, p.404 and Proposition 2.41],

$$\mathbb{P}(\eta\{h(\theta_0) + e_{\theta_0}(X_1)\} \in \mathrm{Exp}_{\theta_0}^{-1}(\mathsf{A})) = \int_{\mathrm{Exp}_{\theta_0}^{-1}(\mathsf{A})} \phi\left(\eta^{-1} v - h(\theta_0)\right) \mathrm{dLeb}_{\theta_0}(v)$$
$$= \int_{\mathsf{A}} \phi\left(\eta^{-1} \mathrm{Exp}_{\theta_0}^{-1}(\theta) - h(\theta_0)\right) \{\det(\mathfrak{g}_{ij}(\theta))\}^{-1/2} \, \mathrm{d}\mu_{\mathsf{S}}(\theta) > 0 ,$$

since all quantities in the integral are positive and $\mu_{\mathsf{S}}(\mathsf{A}) > 0$.

Now assume **A**1-(ii) and keep the notations of the first case. Then $\mathrm{Exp}_{\theta_0} : \mathrm{T}_{\theta_0}\Theta \to \Theta$ is no longer a diffeomorphism. However, $(\mathrm{Exp}_{\theta_0})_{|\mathrm{ID}(\theta_0)} : \mathrm{ID}(\theta_0) \to \Theta \setminus \mathrm{Cut}(\theta_0)$ is a diffeomorphism, see [1, Theorem 10.34]. Moreover, as $\mathrm{Cut}(\theta_0)$ is a set of measure zero, see again [1, Theorem 10.34], considering $\tilde{\mathsf{A}} = \mathsf{A} \setminus \mathrm{Cut}(\theta_0)$ allows the previous proof to give the desired result. □

Proof of Theorem 2. First, we prove that the chain is Harris-recurrent. For that, we start by proving, for any $\theta_0 \in \mathsf{S}$,

$$\mathbb{P}\left(\cup_{k \in \mathbb{N}^*} \cap_{N \in \mathbb{N}} \cup_{n \geq N} \{\theta_n \in \overline{\mathrm{B}}(\theta^\star, k)\}\right) = 1 , \tag{S8}$$

where $(\theta_n)_{n \in \mathbb{N}}$ is defined by (2) and with initial condition θ_0.

Theorem 1-(6) implies that for any $\theta_0 \in \Theta$, $\sup_{n \in \mathbb{N}} Q_\eta^n V(\theta_0) < +\infty$; since $\|V\|_{\mathsf{K}^\star} = \sup_{\mathsf{K}^\star} V < +\infty$ because V is assumed to be continuous. Therefore $\liminf_{n \to +\infty} V(\theta_n)$ is integrable by Fatou's lemma. Thus, for any $k \in \mathbb{N}^*$, using Markov's inequality,

$$\mathbb{P}\left(\liminf_{n \to +\infty} V(\theta_n) > k\right) \leq \mathbb{E}\left[\liminf_{n \to +\infty} V(\theta_n)\right] \Big/ k .$$

However, $\{\liminf_{n \to +\infty} V(\theta_n) \leq k\} = \cap_{N \in \mathbb{N}} \cup_{n \geq N} \{\theta_n \in V^{-1}([0, k])\}$. Thus, for any $k \in \mathbb{N}^*$,

$$\mathbb{P}\left(\cap_{N \in \mathbb{N}} \cup_{n \geq N} \left\{\theta_n \in V^{-1}([0, k])\right\}\right) \geq 1 - \mathbb{E}\left[\liminf_{n \to +\infty} V(\theta_n)\right] \Big/ k .$$

Now, taking the union of these events for any $k \in \mathbb{N}^*$ gives

$$\mathbb{P}\left(\cup_{k \in \mathbb{N}^*} \cap_{N \in \mathbb{N}} \cup_{n \geq N} \left\{\theta_n \in V^{-1}([0, k])\right\}\right) = 1 . \tag{S9}$$

Nonetheless, using **H**1-(iii), for any $k \in \mathbb{N}^*$, $V^{-1}([0, k])$ is a subset of a compact set, therefore it is bounded. Thus, for any $k \in \mathbb{N}^*$, there exists $k' \in \mathbb{N}^*$ such that $V^{-1}([0, k]) \subset \overline{\mathrm{B}}(\theta^\star, k')$. This gives the following,

$$\cup_{k \in \mathbb{N}^*} \cap_{N \in \mathbb{N}} \cup_{n \geq N} \left\{\theta_n \in V^{-1}([0, k])\right\} \subset \cup_{k \in \mathbb{N}^*} \cap_{N \in \mathbb{N}} \cup_{n \geq N} \left\{\theta_n \in \overline{\mathrm{B}}(\theta^\star, k)\right\} .$$

Combining this with (S9) gives (S8).

Equation (S8) gives that the chain is non-evanescent [3, Section 9.2.1]. Since Q_η is Feller (see Lemma S3), this result and [3, Theorem 9.2.2] imply that Q_η is Harris recurrent.

We now show that Q_η is $\tilde{V}$-uniformly geometrically ergodic (see Section S6.1) setting $\tilde{V} = 1 + V$. First, by Theorem 1 and (S7) obtained in the proof above, we have that for any $\theta_0 \in \mathsf{S}, \eta \in (0, \overline{\eta}]$,

$$Q_\eta \tilde{V}(\theta_0) \leq (1 - \eta a)\tilde{V}(\theta_0) + \eta(\eta b/2 + a(1 + \|V\|_{\mathsf{K}^\star})) ,$$

where $a, b, \overline{\eta}$ and $\|V\|_{\mathsf{K}^\star}$ are defined in Theorem 1. Then, by **H**1-(iii) there exists $\tilde{r} > 0$, such that for any $\theta_0 \in \mathsf{S}$,

$$Q_\eta \tilde{V}(\theta_0) \leq (1 - a\eta/2)\tilde{V}(\theta_0) + \eta(\eta b/2 + a(1 + \|V\|_{\mathsf{K}^\star}))\mathbb{1}_{\overline{\mathrm{B}}(\theta^\star, \tilde{r})}(\theta_0) .$$

Then, since Q_η is Feller by Lemma S3 and μ_{S}-irreducible by Lemma S4, using [3, Proposition 6.2.8 (ii)], $\overline{\mathrm{B}}(\theta^\star, r)$ is petite since it is compact by the Hopf-Rinow theorem [4, Theorem 1.7.1] and S has non-empty interior by **A**1. Therefore, an application of [3, Theorem 16.0.1] proves that the chain is $\tilde{V}$-uniformly geometrically ergodic. □

S3.4 Proof of Theorem 3

Lemma S5. *Assume **A**1, **MD**1 **MD**2, **H**1, **H**2 and **H**3($\mathsf{K}^\star$) hold for some compact set $\mathsf{K}^\star \subset \mathsf{S}$. Then for any $\eta \in (0, \overline{\eta}]$,*

$$\mu^\eta[V\mathbb{1}_{\mathsf{S}\setminus\mathsf{K}^\star}] \leq 2\eta L\{\sigma_0^2 + C_1(1 + \sigma_1^2)\}/\lambda ,$$

where $\overline{\eta} = [2C_2L(1 + \sigma_1^2)]^{-1}$.

Proof. For any $\eta \in (0, \overline{\eta}]$ and $M \geq 0$, setting $V_M = M \wedge V$, (S6) implies using Jensen inequality, for any $\theta_0 \in \Theta$,

$$Q_\eta V_M(\theta_0) \leq (1 - \eta a\mathbb{1}_{\mathsf{S}\setminus\mathsf{K}^\star}(\theta_0))V_M(\theta_0) + \eta^2 b/2 ,$$

where $\overline{\eta} = [2C_2L(1 + \sigma_1^2)]^{-1}$, $b = 2L\{\sigma_0^2 + C_1(1 + \sigma_1^2)\}$ and $a = \lambda/2$. Using that μ^η is invariant for Q_η by Theorem 2 and V_M is bounded, we get $\mu^\eta[V_M\mathbb{1}_{\mathsf{S}\setminus\mathsf{K}^\star}] \leq \eta b/(2a)$. By the monotone convergence theorem, taking $M \to +\infty$, we have $\mu^\eta[V\mathbb{1}_{\mathsf{S}\setminus\mathsf{K}^\star}] \leq \eta b/(2a)$, which concludes the proof. □

Proof of Theorem 3. (a) Using Lemma S5 and $V(\theta) \geq c > 0$ for any $\theta \in \mathsf{S} \setminus \mathsf{K}^\star$, we obtain

$$\mu^\eta\{\mathsf{S} \setminus \mathsf{K}^\star\} \leq \eta b/(2ac) ,$$

which concludes the proof of (a) taking the limit $\eta \to 0$.
(b) Let $(\eta_n)_{n\in\mathbb{N}}$ be a sequence converging to zero such that for any $n \in \mathbb{N}$, $\eta_n \in (0, \overline{\eta}]$. We start by proving that $(\mu^{\eta_n})_{n\in\mathbb{N}}$ is tight. Let $\varepsilon > 0$. On one hand, let $r > 0$ and $\mathsf{K}_0 = \overline{\mathrm{B}}(\theta^\star, r)$. Then, using Theorem 3-(a), there exists $N \in \mathbb{N}$ such that for any $n \geq N$, $\mu^{\eta_n}(\mathsf{K}_0) \geq 1 - \varepsilon$. On the other hand, $(\mu^{\eta_n})_{n\in\{0,\ldots,N-1\}}$ is tight, *i.e.* there exists a compact set $\tilde{\mathsf{K}} \subset \Theta$ such that for any $n \in \{1, \ldots, N-1\}$, $\mu^{\eta_n}(\tilde{\mathsf{K}}) \geq 1 - \varepsilon$. Finally, taking $\mathsf{K} = \mathsf{K}_0 \cup \tilde{\mathsf{K}}$ gives the tightness of $(\mu^{\eta_n})_{n\in\mathbb{N}}$. Now, let μ be a limit point of $(\mu^{\eta_n})_{n\in\mathbb{N}}$. Using Theorem 3-(a), and Lebesgue's dominated convergence theorem letting $r \to 0$, gives $\mu(\{\theta^\star\}) = 1$, *i.e.* $\mu = \delta_{\theta^\star}$. In conclusion, for any $(\eta_n)_{n\in\mathbb{N}}$ converging to zero, $(\mu^{\eta_n})_{n\in\mathbb{N}}$ converges weakly to the Dirac at $\theta^\star$.

□

S3.5 Proof of Proposition 4

First, we check **H**1-(i). Using [5, Proposition 2.6], $\mathrm{proj}_{\mathsf{S}}$ is a contraction w.r.t. ρ_Θ, which implies that for any $\theta \in \Theta$,

$$\rho_\Theta^2(\theta^\star, \mathrm{proj}_{\mathsf{S}}(\theta)) = \rho_\Theta^2(\mathrm{proj}_{\mathsf{S}}(\theta^\star), \mathrm{proj}_{\mathsf{S}}(\theta)) \leq \rho_\Theta^2(\theta^\star, \theta) .$$

This implies, since $\mathsf{S} \subset \mathsf{H}$, that

$$V_2(\mathrm{proj}_{\mathsf{S}}(\theta)) = \rho_\Theta^2(\theta^\star, \mathrm{proj}_{\mathsf{S}}(\theta)) \leq \chi_{\mathsf{H}}(\theta)\rho_\Theta^2(\theta^\star, \theta) + (1 - \chi_{\mathsf{H}}(\theta))\,\mathrm{diam}^2(\overline{\mathsf{H}}) = V_2(\theta) ,$$

which gives **H**1-(i).

To prove **H**1-(ii), we calculate the operator norm of the Hessian of V_2 and conclude by [2, Lemma 10]. Using **A**2 and [4, Theorem 5.6.1], $\theta \mapsto \rho_\Theta^2(\theta^\star, \theta)$ is smooth and its gradient on Θ is given by $\theta \mapsto -2\mathrm{Exp}_\theta^{-1}(\theta^\star)$. Therefore, for any $\theta \in \Theta$,

$$\mathrm{grad}\, V_2(\theta) = [\rho_\Theta^2(\theta^\star, \theta) - \mathrm{D}_{\mathsf{H}}^2]\mathrm{grad}\, \chi_{\mathsf{H}}(\theta) - 2\chi_{\mathsf{H}}(\theta)\mathrm{Exp}_\theta^{-1}(\theta^\star) .$$

Using now **A**2, [4, Theorem 5.6.1] and Cauchy-Schwarz's inequality brings, for any $\theta \in \Theta, v \in \mathrm{T}_\theta\Theta$,

$$\|(\mathrm{Hess}\, V_2)_\theta(v, v)\|_\theta \leq 2\kappa\rho_\Theta(\theta^\star, \theta)\coth(\kappa\rho_\Theta(\theta^\star, \theta))\chi_{\mathsf{H}}(\theta)\,\|v\|_\theta^2 + 4\rho_\Theta(\theta^\star, \theta)\,\|\mathrm{grad}\,\chi_{\mathsf{H}}(\theta)\|_\theta\,\|v\|_\theta^2$$

$$+ \|(\mathrm{Hess}\, \chi_{\mathsf{H}})_\theta(v,v)\|_\theta \left|\rho^2_\Theta(\theta^\star,\theta) - \mathrm{D}^2_{\mathsf{H}}\right| .$$

However, one can choose χ_{H} such that for any $\theta \in \Theta$ satisfying $\inf_{\theta' \in \mathsf{H}} \rho_\Theta(\theta',\theta) \geq 1$, it holds that $\chi_{\mathsf{H}}(\theta) = 0$. Therefore, for any $\theta \in \Theta$, $\rho_\Theta(\theta^\star,\theta)\chi_{\mathsf{H}}(\theta) \leq \mathrm{D}_{\mathsf{H}} + 1$. Since χ_{H} is smooth with compact support, there exists a constant $M > 0$ such that for any $\theta \in \Theta$ and $v \in \mathrm{T}_\theta\Theta$,

$$\|\mathrm{grad}\, \chi_{\mathsf{H}}(\theta)\|_\theta \leq M \quad \text{and} \quad \|(\mathrm{Hess}\, \chi_{\mathsf{H}})_\theta(v,v)\|_\theta \leq M \|v\|^2_\theta .$$

Therefore, combining these expressions brings for any $\theta \in \Theta$ and $v \in \mathrm{T}_\theta\Theta$,

$$\|(\mathrm{Hess}\, V_2)_\theta(v,v)\|_\theta \leq 6(M+1)(\mathrm{D}_{\mathsf{H}}+1)[1+\kappa\coth(\kappa\mathrm{D}_{\mathsf{H}})] \|v\|^2_\theta ,$$

thus proving by [2, Lemma 10] and setting $C_\chi = 6(M+1)$, that **H**1-(ii) holds with $L \leftarrow C_\chi(1+\mathrm{D}_{\mathsf{H}})[1+\kappa\coth(\kappa\mathrm{D}_{\mathsf{H}})]$.

We now turn on checking **H**3($\overline{\mathrm{B}}(\theta^\star, r)$). Since $\mathrm{grad}\, \chi_{\mathsf{H}}(\theta) = 0$ for any $\theta \in \mathsf{S}$, we get that V_2 is smooth and for any $\theta \in \mathsf{S}$, $\mathrm{grad}\, V_2(\theta) = -2\mathrm{Exp}^{-1}_\theta(\theta^\star)$ Therefore **H**3($\overline{\mathrm{B}}(\theta^\star, r)$) holds by (8).

S3.6 Proof of Proposition 5

First, we check **H**1-(i). Using [5, Proposition 2.6], $\mathrm{proj}_{\mathsf{S}}$ is a contraction w.r.t. ρ_Θ, which implies that $\theta \in \Theta$,

$$\rho_\Theta(\theta^\star, \mathrm{proj}_{\mathsf{S}}(\theta)) = \rho_\Theta(\mathrm{proj}_{\mathsf{S}}(\theta^\star), \mathrm{proj}_{\mathsf{S}}(\theta)) \leq \rho_\Theta(\theta^\star,\theta) .$$

Then the proof of **H**1-(i) is completed using that $x \mapsto \delta^2\{(x/\delta)^2+1\}^{1/2} - \delta^2$ is increasing.

Next, using **A**2, [2, Lemma 16], we have for any $\theta \in \Theta, v \in \mathrm{T}_\theta\Theta \setminus \{0\}$,

$$0 < \mathrm{Hess}\, V_1(\theta)(v,v) \leq (1+\kappa\delta) \|v\|^2_\theta .$$

Therefore, using [2, Lemma 10], **H**1-(ii) holds for $L = 1+\kappa\delta$. It is easy to see that as $\rho_\Theta(\theta^\star,\theta) \to \infty$, $V_1(\theta) \to +\infty$, meaning **H**1-(iii) holds by the Hopf-Rinow theorem [4, Theorem 1.7.1].

Regarding **H**3($\overline{\mathrm{B}}(\theta^\star, r)$), using [2, Lemma 16], we have for any $\theta \in \Theta$,

$$\mathrm{grad}\, V_1(\theta) = -\, \mathrm{Exp}^{-1}_\theta(\theta^\star) \Big/ \left\{(\rho_\Theta(\theta^\star,\theta)/\delta)^2 + 1\right\}^{1/2} , \tag{S10}$$

Therefore for any $\theta \in \Theta$, we get

$$\langle \mathrm{grad}\, V_1(\theta), h(\theta)\rangle_\theta = -\left\langle \mathrm{Exp}^{-1}_\theta(\theta^\star), h(\theta)\right\rangle_\theta \Big/ \left\{(\rho_\Theta(\theta^\star,\theta)/\delta)^2 + 1\right\}^{1/2} .$$

Then, under the condition (8), we obtain

$$\begin{aligned}\langle \mathrm{grad}\, V_1(\theta), h(\theta)\rangle_\theta &\leq -\lambda_\rho \rho^2_\Theta(\theta^\star,\theta) \mathbb{1}_{\mathsf{S}\setminus\overline{\mathrm{B}}(\theta^\star,r)}(\theta) \Big/ \left\{(\rho_\Theta(\theta^\star,\theta)/\delta)^2 + 1\right\}^{1/2} \\ &\leq -\lambda_\rho V_1(\theta) \mathbb{1}_{\mathsf{S}\setminus\overline{\mathrm{B}}(\theta^\star,r)}(\theta) ,\end{aligned}$$

where we used that

$$V_1(\theta) \leq \rho^2_\Theta(\theta^\star,\theta) \Big/ \left\{(\rho_\Theta(\theta^\star,\theta)/\delta)^2 + 1\right\}^{1/2} ,$$

since for any $a > 0$ and $x \geq 0$, $(ax^2+1)^{1/2} - 1 = a\int_0^x t\{at^2+1\}^{-1/2}\mathrm{d}t \leq ax^2/\{ax^2+1\}^{1/2}$.

S4 Proofs of Section 3

For any $K \in \mathbb{R}_+$, consider a smooth function with compact support $\chi_K : \mathbb{R}_+ \to [0,1]$ such that $\chi_K(t) = 1$ for any $t \leq K$ and $\chi_K(t) = 0$ for any $t \geq K+1$.

Lemma S6. *Assume **A**1-(ii) and **MD**1.*

(a) Then, for any smooth function with compact support $g : \Theta \to \mathbb{R}$, any $\eta > 0$ and $\theta_0 \in \Theta$,

$$Q_\eta g(\theta_0) = g(\theta_0) + \eta \langle \operatorname{grad} g(\theta_0), h(\theta_0)\rangle_{\theta_0} + (\eta^2/2) [\operatorname{Hess} g : \Sigma + h \otimes h] (\theta_0) + (\eta^2/6)\mathscr{R}_{g,\eta}(\theta_0) \,, \tag{S11}$$

where for any $K > 0$,

$$|\mathscr{R}_{g,\eta}(\theta_0)| \le 8\eta\mathbb{E}\left[\|\nabla\operatorname{Hess} g\|_{\gamma,\infty}\mathbb{1}_{\mathsf{A}^{\complement}_{\theta_0}} \|H_K\|^3_{\theta_0}\right] + 16\|\operatorname{Hess} g\|_\infty \mathbb{E}\left[\|Y_K\|^2_{\theta_0}\right] \,, \tag{S12}$$

$$H_K = h(\theta_0) + e_{\theta_0}(X_1)\chi_K(\|e_{\theta_0}(X_1)\|_{\theta_0}) \,, \quad Y_K = e_{\theta_0}(X_1)\{1 - \chi_K(\|e_{\theta_0}(X_1)\|_{\theta_0})\} \,, \tag{S13}$$

$$\|\operatorname{Hess} g\|_\infty = \sup\{|\operatorname{Hess} g_\theta(u, u)| \, : \, \theta \in \Theta, u \in \mathrm{U}_\theta\Theta\} \,,$$

$$\|\nabla\operatorname{Hess} g\|_{\gamma,\infty} = \sup\{|\nabla\operatorname{Hess} g_{\gamma(t)}(u, u, u)| \, : \, t \in [0, 1], u \in \mathrm{U}_{\gamma(t)}\Theta\} \,,$$

$\mathsf{A}_{\theta_0} = \{\|H_K\|_{\theta_0} \le \|Y_K\|_{\theta_0}\}$ *and $\gamma : [0, 1] \to \Theta$ is defined for any $t \in [0, 1]$ by $\gamma(t) = \operatorname{Exp}_{\theta_0}(t\eta H_{\theta_0}(X_1))$.*
(b) Assume in addition that there exist $C_3 > 0$ and $\theta^\star \in \Theta$ such that for any $\theta \in \Theta$, $\|h(\theta)\|_\theta \le C_3\rho_\Theta(\theta^\star, \theta)$. Then, for any smooth function with compact support $g : \Theta \to \mathbb{R}$, any $\eta \in (0, (4C_3)^{-1}]$ and $\theta_0 \in \Theta$, (S11) holds, with for any $K > 0$,

$$|\mathscr{R}_{g,\eta}(\theta_0)| \le 8\eta\mathbb{1}_{\mathsf{K}_K}(\theta_0)\mathbb{E}\left[\|\nabla\operatorname{Hess} g\|_{\gamma,\infty}\mathbb{1}_{\mathsf{A}^{\complement}_{\theta_0}} \|H_K\|^3_{\theta_0}\right] + 16\|\operatorname{Hess} g\|_\infty \mathbb{E}\left[\|Y_K\|^2_{\theta_0}\right] \,, \tag{S14}$$

where we take the notation of (a) and K_K is a compact subset of Θ.

Proof. (a) Let $g : \Theta \to \mathbb{R}$ be a smooth function with compact support and $\theta_0 \in \Theta$. Using (2), **A**1-(ii) and the definition of Q_η (S1), we have

$$Q_\eta g(\theta_0) = \mathbb{E}\left[g\left\{\operatorname{Exp}_{\theta_0}[\eta H_{\theta_0}(X_1)]\right\}\right] \,. \tag{S15}$$

Consider the geodesic $\gamma : [0, 1] \to \Theta$ defined for any $t \in [0, 1]$ by $\gamma(t) = \operatorname{Exp}_{\theta_0}(t\eta H_{\theta_0}(X_1))$. For any $t \in [0, 1]$, let $\mathscr{g}(t) = (g \circ \gamma)(t)$. We compute now its derivatives to derive a Taylor expansion. Using [1, Proposition 4.15-(ii) and Theorem 4.24-(iii)], we have for any $t \in [0, 1]$,

$$\mathscr{g}'(t) = \mathrm{D}_t(g \circ \gamma)(t) = \langle \operatorname{grad} g(\gamma(t)), \dot\gamma(t)\rangle_{\gamma(t)} \,.$$

By definition of the Hessian [1, Example 4.22] and using $\mathrm{D}_t\dot\gamma(t) = 0$, Proposition S14-(S72)-(iv), we get for any $t \in [0, 1]$,

$$\mathscr{g}''(t) = [\mathrm{D}^2_t\mathscr{g}](t) = \operatorname{Hess} g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t)) \,,$$

In addition, using $\mathrm{D}_t\dot\gamma(t) = 0$ and Proposition S14-(S72)-(iv), we obtain for any $t \in [0, 1]$,

$$\mathscr{g}^{(3)}(t) = [\mathrm{D}^3_t\mathscr{g}](t) = \nabla\operatorname{Hess} g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t), \dot\gamma(t)) \,,$$

where $\nabla\operatorname{Hess} g$ is the total covariant derivative of $\operatorname{Hess} g$ [1, Proposition 4.17]. Finally, for any $K > 0$, consider the two random tangent vectors at θ_0 defined in (S13). Now, writing the first-order Taylor expansion of $\mathscr{g} : [0, 1] \to \mathbb{R}$, at $t = 1$ on the event $\mathsf{A}_{\theta_0} = \{\|H_K\|_{\theta_0} \le \|Y_K\|_{\theta_0}\}$, the second-order one on the complement, and summing both expansions, we get

$$\begin{aligned} g\left(\operatorname{Exp}_{\theta_0}(\eta H_{\theta_0}(X_1))\right) = g(\theta_0) &+ \eta\langle \operatorname{grad} g(\theta_0), H_{\theta_0}(X_1)\rangle_{\theta_0} \\ &+ (\eta^2/2)\operatorname{Hess} g_{\theta_0}(H_{\theta_0}(X_1), H_{\theta_0}(X_1)) + \mathscr{R}_{g,\eta}(\theta_0, X_1)/6 \,, \end{aligned} \tag{S16}$$

where the remainder term is given by

$$\begin{aligned} \mathscr{R}_{g,\eta}(\theta_0, X_1) = \mathbb{1}_{\mathsf{A}^{\complement}_{\theta_0}} &\int_0^1 \nabla\operatorname{Hess} g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t), \dot\gamma(t))\mathrm{d}t \\ &+ \mathbb{1}_{\mathsf{A}_{\theta_0}}\left[\int_0^1 \operatorname{Hess} g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t))\mathrm{d}t - 3\eta^2\operatorname{Hess} g_{\theta_0}(H_{\theta_0}(X_1), H_{\theta_0}(X_1))\right] \,. \end{aligned}$$

We bound the remainder as follows. Since g has compact support, $\operatorname{Hess} g$ and $\nabla\operatorname{Hess} g$ have an operator norm uniformly bounded over Θ, which we express in the following way. For any $\theta \in \Theta$, consider the unit tangent space at θ, $\mathrm{U}_\theta\Theta = \{v \in \mathrm{T}_\theta\Theta \, : \, \|v\|_\theta = 1\}$, let $\|\operatorname{Hess} g\|_\infty = \sup\{|\operatorname{Hess} g_\theta(v, v)| \, : \, \theta \in \Theta, v \in \mathrm{U}_\theta\Theta\}$ and

$\|\nabla \text{Hess}\, g\|_{\gamma,\infty} = \sup\{|\nabla \text{Hess}\, g_{\gamma(t)}(v,v,v)| : t \in [0,1], v \in \text{U}_{\gamma(t)}\Theta\}$. Then, using [1, Corollary 5.6-(b)], and $\dot{\gamma}(0) = \eta H_{\theta_0}(X_1)$,

$$\begin{aligned} |\mathscr{R}_{g,\eta}(\theta_0, X_1)| &\leq \mathbb{1}_{\mathsf{A}^{\complement}_{\theta_0}} \|\nabla \text{Hess}\, g\|_{\gamma,\infty} \int_0^1 \|\dot{\gamma}(t)\|^3_{\gamma(t)}\, \mathrm{d}t \\ &\quad + \mathbb{1}_{\mathsf{A}_{\theta_0}} \|\text{Hess}\, g\|_\infty \left[\int_0^1 \|\dot{\gamma}(t)\|^2_{\gamma(t)}\, \mathrm{d}t + 3\eta^2 \|H_{\theta_0}(X_1)\|^2_{\theta_0}\right] \\ &= \mathbb{1}_{\mathsf{A}^{\complement}_{\theta_0}} \|\nabla \text{Hess}\, g\|_{\gamma,\infty} \eta^3 \|H_{\theta_0}(X_1)\|^3_{\theta_0} + 4\mathbb{1}_{\mathsf{A}_{\theta_0}} \|\text{Hess}\, g\|_\infty \eta^2 \|H_{\theta_0}(X_1)\|^2_{\theta_0} \ . \end{aligned}$$

Moreover, using that $H_K + Y_K = H_{\theta_0}(X_1)$ and the definition of A_{θ_0},

$$|\mathscr{R}_{g,\eta}(\theta_0, X_1)| \leq 8\mathbb{1}_{\mathsf{A}^{\complement}_{\theta_0}} \|\nabla \text{Hess}\, g\|_{\gamma,\infty} \eta^3 \|H_K\|^3_{\theta_0} + 16\|\text{Hess}\, g\|_\infty \eta^2 \|Y_K\|^2_{\theta_0} \ . \tag{S17}$$

Now, using **MD1**,

$$\mathbb{E}\left[\langle \text{grad}\, g(\theta_0), H_{\theta_0}(X_1)\rangle_{\theta_0}\right] = \langle \text{grad}\, g(\theta_0), h(\theta_0)\rangle_{\theta_0} \ . \tag{S18}$$

In addition, since

$$\text{Hess}\, g_{\theta_0}(H_{\theta_0}(X_1), H_{\theta_0}(X_1)) = \left[\text{Hess}\, g : H_{\theta_0}(X_1) \otimes H_{\theta_0}(X_1)\right] \ ,$$

it follows by a further application of **MD1**, that

$$\mathbb{E}\left[\text{Hess}\, g_{\theta_0}(H_{\theta_0}(X_1), H_{\theta_0}(X_1))\right] = \left[\text{Hess}\, g : h \otimes h + \Sigma\right](\theta_0) \ , \tag{S19}$$

where $\Sigma(\theta_0)$ is defined in (10). Using that $\|H_K\|_{\theta_0} \leq K + \|h(\theta_0)\|_{\theta_0}$, and **MD1** in (S17), we obtain that for any $\theta_0 \in \Theta, \mathbb{E}[|\mathscr{R}_{g,\eta}(\theta_0, X_1)|] < +\infty$. Then, by (S16), (S18) and (S19), it follows from (S15),

$$Q_\eta g(\theta_0) = g(\theta_0) + \eta\langle \text{grad}\, g(\theta_0), h(\theta_0)\rangle_{\theta_0} + (\eta^2/2)\left[\text{Hess}\, g : h \otimes h + \Sigma\right](\theta_0) + \eta^2 \mathscr{R}_{g,\eta}(\theta_0)/6 \ ,$$

where we define $\mathscr{R}_{g,\eta}(\theta_0) = \eta^{-2}\mathbb{E}[\mathscr{R}_{g,\eta}(\theta_0, X_1)]$. The desired bound on the remainder in (S12), is a simple consequence of (S17).

(b) In addition to the results of (a) and specifically (S12), we need to prove that, since g has compact support, there exists a compact set $\mathsf{K}_K \subset \Theta$ such that $\|\nabla \text{Hess}\, g\|_{\gamma,\infty} \mathbb{1}_{\mathsf{A}^{\complement}_{\theta_0}} = 0$ for any $\theta_0 \notin \mathsf{K}_K$.

Using that $\|h(\theta)\|_\theta \leq C_3 \rho_\Theta(\theta^\star, \theta)$, we obtain that on $\mathsf{A}^{\complement}_{\theta_0}$, $\|H_\theta(X_1)\|_\theta \leq 2(C_3 \rho_\Theta(\theta^\star, \theta) + K)$. In addition, by [1, Corollary 6.12], $\rho_\Theta(\theta, \gamma(t)) = t\eta \|H_\theta(X_1)\|_\theta$ for any $t \in [0,1]$, therefore for any $t \in [0,1]$ and $\eta \in (0, (4C_3)^{-1}]$

$$\rho_\Theta(\theta^\star, \gamma(t)) \geq \rho_\Theta(\theta^\star, \theta) - \rho_\Theta(\theta, \gamma(t)) \geq (1 - 2\eta t C_3)\rho_\Theta(\theta^\star, \theta) - 2\eta K \geq \rho_\Theta(\theta^\star, \theta)/2 - K/(2C_3) \ .$$

Consider now $R \geq 0$ such that for any $\theta \notin \overline{\mathrm{B}}(\theta^\star, R)$, $g(\theta) = 0$. Then, setting $\mathsf{K}_K = \overline{\mathrm{B}}(\theta^\star, 2(R + K/(2C_3)))$, we obtain that for any $\theta_0 \notin \mathsf{K}_K$ and $t \in [0,1]$, $\gamma(t) \notin \overline{\mathrm{B}}(\theta^\star, R)$ and therefore, $\nabla \text{Hess}\, g_{\gamma(t)} = 0$, which yields $\|\nabla \text{Hess}\, g\|_{\gamma,\infty} \mathbb{1}_{\mathsf{A}^{\complement}_{\theta_0}} = 0$ for any $\theta_0 \notin \mathsf{K}_K$. Finally K_K is a compact subset of Θ by [4, Theorem 1.7.1].

□

S4.1 Proof of Theorem 6

Let $g : \Theta \to \mathbb{R}$ be a smooth function. Since we assume that Θ is compact, g is smooth with compact support. Therefore, using Lemma S6-(a) for any $\theta \in \Theta$ and $\eta > 0$, we have,

$$Q_\eta g(\theta) = g(\theta) + \eta \langle \text{grad}\, g(\theta), h(\theta)\rangle_\theta + (\eta^2/2)[\text{Hess}\, g : \Sigma + h \otimes h](\theta) + (\eta^2/6)\mathscr{R}_{g,\eta}(\theta) \ , \tag{S20}$$

where using (S12), Hölder inequality and **MD1** gives,

$$\begin{aligned} |\mathscr{R}_{g,\eta}(\theta)| \leq 32\eta(\|h(\theta)\|^3_\theta + K^3) \sup\{|\nabla \text{Hess}\, g_\theta(u,u,u)| : \theta \in \Theta, u \in \text{U}_\theta\Theta\} \\ + 16\, \|\text{Hess}\, g\|_\infty \left(\sigma_0^2 + \sigma_1^2 \|h(\theta)\|^2_\theta\right) \ . \end{aligned}$$

Next, let $\eta \in (0, \bar{\eta}]$, where $\bar{\eta} = [2C_2L(1+\sigma_1^2)]^{-1}$. Note that since Θ is compact, g is smooth, h and Σ are continuous, all the functions appearing in (S20) are bounded. Therefore, integrating (S20) with respect to μ^η given by Theorem 2 and using that μ^η is invariant w.r.t. Q_η, we obtain,

$$-\int_\Theta \langle \operatorname{grad} g(\theta), h(\theta)\rangle_\theta \, \mu^\eta(\mathrm{d}\theta) = (\eta/2)\int_\Theta [\operatorname{Hess} g : \Sigma + h \otimes h](\theta)\mu^\eta(\mathrm{d}\theta) + (\eta/6)\int_\Theta \mathscr{R}_{g,\eta}(\theta)\mu^\eta(\mathrm{d}\theta) \ .$$

Using that $\theta \mapsto [\operatorname{Hess} g : \Sigma + h \otimes h](\theta)$ is bounded and continuous over Θ, Theorem 3-(b) and that $h(\theta^\star) = 0$, by weak convergence of μ^η to $\delta_{\theta^\star}$ when $\eta \to 0$, we have,

$$\lim_{\eta\to 0}\int_\Theta [\operatorname{Hess} g : \Sigma + h \otimes h](\theta)\mu^\eta(\mathrm{d}\theta) = [\operatorname{Hess} g : \Sigma + h \otimes h](\theta^\star) = [\operatorname{Hess} g : \Sigma](\theta^\star) \ .$$

Equivalently, there exists $\mathscr{R}_{\operatorname{Hess} g} : (0, \bar{\eta}] \to \mathbb{R}$ such that for any $\eta \in (0, \bar{\eta}]$, we have

$$\int_\Theta [\operatorname{Hess} g : \Sigma + h \otimes h](\theta)\mu^\eta(\mathrm{d}\theta) = [\operatorname{Hess} g : \Sigma](\theta^\star) + \mathscr{R}_{\operatorname{Hess} g}(\eta) \ ,$$

where $\lim_{\eta\to 0} |\mathscr{R}_{\operatorname{Hess} g}(\eta)| = 0$.

To conclude, we prove that $\limsup_{\eta\to 0} |\int_\Theta \mathscr{R}_{g,\eta}(\theta)\mu^\eta(\mathrm{d}\theta)| = 0$. Let $K \geq 0$. By (S12), since $\theta_0 \mapsto \mathbb{E}[\mathbb{1}_{\mathsf{A}^{\complement}_{\theta_0}} \|H_K\|^3_{\theta_0}]$ is uniformly bounded over Θ by definition (S13) and since h is continuous, we have that

$$\begin{aligned}\limsup_{\eta\to 0}\left|\int_\Theta \mathscr{R}_{g,\eta}(\theta)\mu^\eta(\mathrm{d}\theta)\right| &\leq 16\|\operatorname{Hess} g\|_\infty \limsup_{\eta\to 0}\int_\Theta \mathbb{E}\left[\|e_\theta(X_1)\|^2_\theta \{1 - \chi_K(\theta)\}\right]\mu^\eta(\mathrm{d}\theta) \\ &\leq 16\|\operatorname{Hess} g\|_\infty \mathbb{E}\left[\|e_{\theta^\star}(X_1)\|^2_{\theta^\star}\{1 - \chi_K(\theta^\star)\}\right] \ ,\end{aligned}$$

using Theorem 3-(b), that $\theta \mapsto \mathbb{E}[\|e_\theta(X_1)\|^2_\theta]$ and χ_K are continuous and bounded by **MD**3 since $\mathbb{E}[\|e_\theta(X_1)\|^2_\theta] = \operatorname{Tr}(\Sigma(\theta))$ for any $\theta \in \Theta$ and Θ is compact. Taking $K \to +\infty$ completes the proof.

S4.2 Proof of Theorem 7

We introduce an auxiliary chain $(U_n)_{n\in\mathbb{N}}$ as an intermediate step between $(\theta_n)_{n\in\mathbb{N}}$ and $(\overline{U}_n)_{n\in\mathbb{N}}$ for which we recall the definition below. Define for any $\eta > 0, n \in \mathbb{N}$,

$$U_n = \operatorname{Exp}^{-1}_{\theta^\star}(\theta_n) \quad \text{and} \quad \overline{U}_n = \eta^{-1/2}\operatorname{Exp}^{-1}_{\theta^\star}(\theta_n) = \eta^{-1/2}U_n \ , \tag{S21}$$

where $(\theta_n)_{n\in\mathbb{N}}$ is defined by (2) with $\mathsf{S} = \Theta$ *i.e.* $\operatorname{proj}_{\mathsf{S}} = \operatorname{Id}$. Note that $(U_n)_{n\in\mathbb{N}}$ and $(\overline{U}_n)_{n\in\mathbb{N}}$ are Markov chains with state space $\mathrm{T}_{\theta^\star}\Theta$, as $\operatorname{Exp}_{\theta^\star}$ is a bijection. Conversely, since $\operatorname{Exp}^{-1}_{\theta^\star}$ and $\eta^{-1/2}\operatorname{Exp}^{-1}_{\theta^\star}$ are bijections from Θ to $\mathrm{T}_{\theta^\star}\Theta$ under **A**1-(i), $(\theta_n)_{n\in\mathbb{N}}$ is a deterministic function of $(U_n)_{n\in\mathbb{N}}$ or $(\overline{U}_n)_{n\in\mathbb{N}}$. Therefore, the convergence of these three processes is expected to be the same. This is the content of the following result. Denote by R_η and $\overline{R}_\eta$ the Markov kernels on $\mathrm{T}_{\theta^\star}\Theta \times \mathcal{B}(\mathrm{T}_{\theta^\star}\Theta)$, associated with $(U_n)_{n\in\mathbb{N}}$ and $(\overline{U}_n)_{n\in\mathbb{N}}$ respectively.

Lemma S7. *Assume **A**1-(i)-(ii), **MD**1, **MD**2, **H**1, **H**2 and **H**3($\mathsf{K}^\star$) for some compact set $\mathsf{K}^\star \subset \mathsf{S}$. Let $\eta \in (0, \bar{\eta}]$ where $\bar{\eta} = [2C_2L(1+\sigma_1^2)]^{-1}$. For any measurable and bounded function $\mathrm{g} : \mathrm{T}_{\theta^\star}\Theta \to \mathbb{R}$ and any $u_0, \bar{u}_0 \in \mathrm{T}_{\theta^\star}\Theta$, R_η and $\overline{R}_\eta$ satisfy*

$$R_\eta \mathrm{g}(u_0) = Q_\eta g\left(\operatorname{Exp}_{\theta^\star}(u_0)\right) \quad \text{and} \quad \overline{R}_\eta \mathrm{g}\left(\bar{u}_0\right) = R_\eta \mathrm{g}_\eta(\eta^{1/2}\bar{u}_0) \ , \tag{S22}$$

where $g : \theta \mapsto \mathrm{g}[\operatorname{Exp}^{-1}_{\theta^\star}(\theta)]$ and $\mathrm{g}_\eta : u \mapsto \mathrm{g}(\eta^{-1/2}u)$ are defined over Θ and $\mathrm{T}_{\theta^\star}\Theta$ respectively, and Q_η is the Markov kernel associated with $(\theta_n)_{n\in\mathbb{N}}$. In addition, R_η and $\overline{R}_\eta$ both admit a unique stationary distribution ν^η and $\bar{\nu}^\eta$ respectively, defined for any $\mathsf{A} \in \mathcal{B}(\mathrm{T}_{\theta^\star}\Theta)$ by

$$\nu^\eta(\mathsf{A}) = \mu^\eta\left(\operatorname{Exp}_{\theta^\star}(\mathsf{A})\right) \quad \text{and} \quad \bar{\nu}^\eta(\mathsf{A}) = \nu^\eta(\eta^{1/2}\mathsf{A}) \ . \tag{S23}$$

Finally, both R_η and $\overline{R}_\eta$ are Harris-recurrent and geometrically ergodic, i.e. there exist $C, \overline{C} : \mathrm{T}_{\theta^\star}\Theta \to \mathbb{R}$ and $\rho, \bar{\rho} \in \mathbb{R}^*_+$ such that for any $u, \bar{u} \in \mathrm{T}_{\theta^\star}\Theta$,

$$\|\delta_u R_\eta - \nu^\eta\|_{\mathrm{TV}} \leq C(u)\rho^n \quad \text{and} \quad \|\delta_{\bar{u}}\overline{R}_\eta - \bar{\nu}^\eta\|_{\mathrm{TV}} \leq \overline{C}(\bar{u})\bar{\rho}^n \ .$$

Proof. Let $\mathrm{g} : \mathrm{T}_{\theta^\star}\Theta \to \mathbb{R}$ be a measurable and bounded function and $u_0 \in \mathrm{T}_{\theta^\star}\Theta$. Consider $(U_n)_{n\in\mathbb{N}}$ defined by (S21) with $\theta_0 = \mathrm{Exp}_{\theta^\star}(u_0)$. Using (S21), we have by definition

$$\mathbb{E}\left[\mathrm{g}(U_1)\right] = \mathbb{E}\left[\mathrm{g}\left(\mathrm{Exp}_{\theta^\star}^{-1}(\theta_1)\right)\right] = Q_\eta\left(\mathrm{g}\circ\mathrm{Exp}_{\theta^\star}^{-1}\right)\left(\mathrm{Exp}_{\theta^\star}(u_0)\right) .$$

Moreover, let $\overline{u}_0 \in \mathrm{T}_{\theta^\star}\Theta$ and consider $(\overline{U}_n)_{n\in\mathbb{N}}$ defined by (S21) with $U_0 = \eta^{1/2}\overline{u}_0$. Using (S21), we have by definition

$$\mathbb{E}\left[\mathrm{g}(\overline{U}_1)\right] = \mathbb{E}\left[\mathrm{g}\left(\eta^{-1/2}U_1\right)\right] = R_\eta \mathrm{g}_\eta\left(\eta^{1/2}\overline{u}_0\right) ,$$

where $\mathrm{g}_\eta : u \mapsto \mathrm{g}(\eta^{-1/2}u)$ is defined over $\mathrm{T}_{\theta^\star}\Theta$, therefore proving (S22).

We show that ν^η and $\overline{\nu}^\eta$ are invariant for R_η and $\overline{R}_\eta$ respectively. Indeed, for any $\mathsf{A} \in \mathcal{B}(\mathrm{T}_{\theta^\star}\Theta)$, we have by (S21), (S22) and (S23)

$$\begin{aligned}\nu^\eta R_\eta(\mathsf{A}) &= \int_{\mathrm{T}_{\theta^\star}\Theta} \mathrm{d}\nu^\eta(u) R_\eta\left(u, \mathsf{A}\right) = \int_\Theta \mathrm{d}\mu^\eta(\theta) R_\eta\left(\mathrm{Exp}_{\theta^\star}^{-1}(\theta), \mathsf{A}\right) \\ &= \int_\Theta \mathrm{d}\mu^\eta(\theta) Q_\eta\left(\theta, \mathrm{Exp}_{\theta^\star}(\mathsf{A})\right) = \mu^\eta\left(\mathrm{Exp}_{\theta^\star}(\mathsf{A})\right) = \nu^\eta(A) .\end{aligned}$$

Therefore ν^η is invariant for R_η. Similarly, we show that $\overline{\nu}^\eta$ is invariant for $\overline{R}_\eta$. Using again (S21), (S22) and (S23), for any $\mathsf{A} \in \mathcal{B}(\mathrm{T}_{\theta^\star}\Theta)$ we have,

$$\overline{\nu}^\eta \overline{R}_\eta(\mathsf{A}) = \int_{\mathrm{T}_{\theta^\star}\Theta} \mathrm{d}\nu^\eta(u) \overline{R}_\eta\left(\eta^{-1/2}u, \mathsf{A}\right) = \int_{\mathrm{T}_{\theta^\star}\Theta} \mathrm{d}\nu^\eta(u) R_\eta\left(u, \eta^{1/2}\mathsf{A}\right) = \overline{\nu}^\eta(\mathsf{A}) .$$

Finally, since $(\theta_n)_{n\in\mathbb{N}}$, $(U_n)_{n\in\mathbb{N}}$ and $(\overline{U}_n)_{n\in\mathbb{N}}$ are deterministic functions of each other and since Theorem 2 proves that $(\theta_n)_{n\in\mathbb{N}}$ is geometrically ergodic and Harris-recurrent, the same holds for $(U_n)_{n\in\mathbb{N}}$ and $(\overline{U}_n)_{n\in\mathbb{N}}$ and their invariant distributions are unique. □

For any smooth function with compact support $\mathrm{g} : \mathrm{T}_{\theta^\star}\Theta \to \mathbb{R}$, $\overline{u}_0 \in \mathrm{T}_{\theta^\star}\Theta$ and $\eta > 0$ consider the 2-tensor $(\mathsf{C}^2(\mathrm{g}, \overline{u}_0, \eta)_{ij})_{i,j\in\{1,\dots,d\}}$ defined by, for any $i, j \in \{1, \dots, d\}$,

$$\mathsf{C}^2(\mathrm{g}, \overline{u}_0, \eta)_{ij} = \partial^2_{ij}\mathrm{g}(\overline{u}_0) - \eta^{1/2}\sum_{k=1}^d \Gamma^k_{ij}\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)\right)\partial_k \mathrm{g}(\overline{u}_0) , \tag{S24}$$

and, similarly consider the 3-tensor $(\mathsf{C}^3(\mathrm{g}, \overline{u}_0, \eta)_{ijk})_{i,j,k\in\{1,\dots,d\}}$ defined by, for any $i, j, k \in \{1, \dots, d\}$,

$$\begin{aligned}\mathsf{C}^3(\mathrm{g}, \overline{u}_0, \eta)_{ijk} = \partial^3_{ijk}\mathrm{g}(\overline{u}_0) \\ &- \eta^{1/2}\sum_{l=1}^d \Big[\Gamma^l_{ij}\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)\right)\partial^2_{kl}\mathrm{g}(\overline{u}_0) + \Gamma^l_{ki}\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)\right)\partial^2_{jl}\mathrm{g}(\overline{u}_0) \\ &+ \Gamma^l_{kj}\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)\right)\partial^2_{il}\mathrm{g}(\overline{u}_0)\Big] - \eta\sum_{m=1}^d \partial_k\Gamma^m_{ij}\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)\right)\partial_m\mathrm{g}(\overline{u}_0) \\ &+ \eta\sum_{l,m=1}^d \left[\Gamma^l_{kj}\Gamma^m_{il} + \Gamma^l_{ki}\Gamma^m_{lj}\right]\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)\right)\partial_m\mathrm{g}(\overline{u}_0) ,\end{aligned} \tag{S25}$$

where $(\Gamma^k_{ij})_{i,j,k\in\{1,\dots,d\}}$ are the Christoffel symbols of the Levi-Civita connection ∇. We derive the following Taylor formulas.

Lemma S8. *Assume **A**1-(i)-(ii), **MD**1, **MD**2, **H**1, **H**2 and **H**3($\mathsf{K}^\star$) for some compact set $\mathsf{K}^\star \subset \mathsf{S}$. Suppose in addition that there exists $C_3 > 0$ such that for any $\theta \in \Theta$, $\|h(\theta)\|_\theta \le C_3\rho_\Theta(\theta^\star, \theta)$ and let $\overline{\eta} = [2C_2L(1 + \sigma_1^2)]^{-1} \wedge (4C_3)^{-1}$. Consider normal coordinates $(u^i)_{i\in\{1,\dots,d\}}$ centered at $\theta^\star$ and define for any $i, j \in \{1, \dots, d\}$, $h^i : \Theta \to \mathbb{R}$, $\Sigma_{ij} : \Theta \to \mathbb{R}$ by $h^i = \mathrm{d}u^i(h)$ and $\Sigma_{ij} = [\mathrm{d}u^i \otimes \mathrm{d}u^j]\{\Sigma\}$. For any smooth function with compact support $\mathrm{g} : \mathrm{T}_{\theta^\star}\Theta \to \mathbb{R}$, any $\eta \in (0, \overline{\eta}]$ and $\overline{u}_0 \in \mathrm{T}_{\theta^\star}\Theta$, we have*

$$\overline{R}_\eta \mathrm{g}(\overline{u}_0) = \mathrm{g}(\overline{u}_0) + \eta^{1/2}\sum_{i=1}^d \partial_i\mathrm{g}(\overline{u}_0) h^i\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)\right) \tag{S26}$$

$$+\frac{\eta}{2}\sum_{i,j=1}^{d}\left\{\partial^2_{ij}\mathrm{g}\left(\overline{u}_0\right)-\eta^{1/2}\sum_{k=1}^{d}\Gamma^k_{ij}(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0))\partial_k\mathrm{g}\left(\overline{u}_0\right)\right\}\left[\Sigma_{ij}+h^ih^j\right]\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)\right)$$
$$+(\eta/6)\overline{\mathscr{R}}_{\mathrm{g},\eta}(\overline{u}_0)\;,$$

where, setting $\theta_0=\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)$*,*

$$\left|\overline{\mathscr{R}}_{\mathrm{g},\eta}(\overline{u}_0)\right|\le 8\eta^{1/2}\mathbb{1}_{\mathsf{K}_K}(\theta_0)\mathbb{E}\left[\|\mathsf{C}^3(\mathrm{g},\eta)\|_\gamma\mathbb{1}_{\mathsf{A}^\complement_{\theta_0}}\|H_K\|^3_{\theta_0}\right]+16\|\mathsf{C}^2(\mathrm{g},\eta)\|\mathbb{E}\left[\|Y_K\|^2_{\theta_0}\right]\;,\tag{S27}$$

using the definitions of H_K, Y_K*,* A_{θ_0}*,* K_K *and* γ *in Lemma S6-(S13),*

$$\begin{aligned}\|\mathsf{C}^2(\mathrm{g},\eta)\|&=\sup\{|\mathsf{C}^2(\mathrm{g},\overline{u},\eta)[v^{\otimes 2}]|\,:\,\overline{u}\in\mathrm{T}_{\theta^\star}\Theta, v\in\mathbb{R}^d,\|v\|_2=1\}\\ \|\mathsf{C}^3(\mathrm{g},\eta)\|_\gamma&=\sup\{|\mathsf{C}^3(\mathrm{g},\overline{u},\eta)[v^{\otimes 3}]|\,:\,\overline{u}\in\eta^{-1/2}\mathrm{Exp}^{-1}_{\theta^\star}(\gamma([0,1])), v\in\mathbb{R}^d,\|v\|_2=1\}\;,\end{aligned}\tag{S28}$$

where $\mathsf{C}^2(\mathrm{g},\overline{u},\eta)$ *and* $\mathsf{C}^3(\mathrm{g},\overline{u},\eta)$ *are defined in (S24) and (S25).*

Proof. Using **A**1-(i) and [1, Proposition 12.9], $(u^i)_{i\in\{1,\ldots,d\}}$ are global coordinates on the Hadamard manifold Θ. Let $\mathrm{g}:\mathrm{T}_{\theta^\star}\Theta\to\mathbb{R}$ be a smooth function with compact support and $g:\Theta\to\mathbb{R}$ defined for any $\theta\in\Theta$ by $g(\theta)=\mathrm{g}(\mathrm{Exp}^{-1}_{\theta^\star}(\theta))$. Note that since $\|\mathrm{Exp}^{-1}_{\theta^\star}(\theta)\|_{\theta^\star}=\rho_\Theta(\theta^\star,\theta)$, for any $\theta\in\Theta$ by [1, Corollary 6.12], g is a smooth function with compact support as well. In addition, by definition of the normal coordinates, $\mathrm{g}:u\mapsto g(\mathrm{Exp}_{\theta^\star}(u))$ is the expression of g in this coordinate system. Using this fact and the definitions of the Riemannian gradient and Hessian [1, Equation 2.14, Example 4.22], we have, for any $\theta_0\in\Theta$,

$$\begin{aligned}\mathrm{grad}\,g(\theta_0)&=\sum_{i=1}^d\partial_i\mathrm{g}(u_0)\partial u_i\;,\\ \mathrm{Hess}\,g(\theta_0)&=\sum_{i,j=1}^d\left\{\partial^2_{ij}\mathrm{g}(u_0)-\sum_{k=1}^d\Gamma^k_{ij}(\mathrm{Exp}_{\theta^\star}(u_0))\partial_k\mathrm{g}(u_0)\right\}\mathrm{d}u^i\otimes\mathrm{d}u^j\;,\end{aligned}\tag{S29}$$

where $u_0=\mathrm{Exp}^{-1}_{\theta^\star}(\theta_0)$ and $(\Gamma^k_{ij})_{i,j,k\in\{1,\ldots,d\}}$ are the Christoffel symbols. Combining these expressions with Lemma S7-(S22) and Lemma S6-(b)-(S11) gives

$$\begin{aligned}R_\eta\mathrm{g}(u)&=\mathrm{g}(u_0)+\eta\sum_{i=1}^d\partial_i\mathrm{g}(u_0)h^i(\mathrm{Exp}_{\theta^\star}(u_0))\\&+(\eta^2/2)\sum_{i,j=1}^d\left\{\partial^2_{ij}\mathrm{g}(u_0)-\sum_{k=1}^d\Gamma^k_{ij}(\mathrm{Exp}_{\theta^\star}(u_0))\partial_k\mathrm{g}(u_0)\right\}\left[\Sigma_{ij}\left(\mathrm{Exp}_{\theta^\star}(u_0)\right)+h^ih^j\left(\mathrm{Exp}_{\theta^\star}(u_0)\right)\right]\\&+(\eta^2/6)\tilde{\mathscr{R}}_{\mathrm{g},\eta}(u_0)\;,\end{aligned}$$

where $\tilde{\mathscr{R}}_{\mathrm{g},\eta}(u_0)=\mathscr{R}_{g,\eta}(\theta_0)$ is bounded using (S14), for $\theta_0=\mathrm{Exp}_{\theta^\star}(u_0)$ and $g:\theta\mapsto\mathrm{g}(\mathrm{Exp}^{-1}_{\theta^\star}(\theta))$.

Replacing g with $\mathrm{g}_\eta:u\mapsto\mathrm{g}(\eta^{-1/2}u)$ defined over $\mathrm{T}_{\theta^\star}\Theta$ and using that for any $i,j\in\{1,\ldots,d\}$ and $u_0\in\mathrm{T}_{\theta^\star}\Theta$,

$$\partial_i\mathrm{g}_\eta(u_0)=\eta^{-1/2}\partial_i\mathrm{g}(\eta^{-1/2}u_0)\quad\text{ and }\quad\partial^2_{ij}\mathrm{g}_\eta(u_0)=\eta^{-1}\partial_{ij}\mathrm{g}(\eta^{-1/2}u_0)\;,\tag{S30}$$

we have for any $u_0\in\mathrm{T}_{\theta^\star}\Theta$,

$$\begin{aligned}R_\eta\mathrm{g}_\eta(u_0)&=\mathrm{g}(\eta^{-1/2}u_0)+\eta^{1/2}\sum_{i=1}^d\partial_i\mathrm{g}(\eta^{-1/2}u_0)h^i\left(\mathrm{Exp}_{\theta^\star}(u_0)\right)\\&+(\eta/2)\sum_{i,j=1}^d\left\{\partial^2_{ij}\mathrm{g}\left(\frac{u_0}{\eta^{1/2}}\right)-\eta^{1/2}\sum_{k=1}^d\Gamma^k_{ij}(\mathrm{Exp}_{\theta^\star}(u_0))\partial_k\mathrm{g}\left(\frac{u_0}{\eta^{1/2}}\right)\right\}\left[\Sigma_{ij}+h^ih^j\right]\left(\mathrm{Exp}_{\theta^\star}(u_0)\right)\\&+(\eta^2/6)\tilde{\mathscr{R}}_{\mathrm{g}_\eta,\eta}(u_0)\;.\end{aligned}\tag{S31}$$

Expressing $\tilde{\mathscr{R}}_{\mathrm{g}_\eta,\eta}(u_0)$ using partial derivatives shows explicitly the dependency on η. Using (S30) and the equivalent formula for the third order derivative, we have for any $K>0$,

$$\eta^2 \left|\tilde{\mathscr{R}}_{\mathrm{g}_\eta,\eta}(u_0)\right| \leq 8\eta^3 \mathbb{1}_{\mathsf{K}_K}(\theta_0)\mathbb{E}\left[\|\nabla \mathrm{Hess}\,\mathrm{g}_\eta\|_{\gamma,\infty}\mathbb{1}_{\mathsf{A}^{\complement}_{\theta_0}}\|H_K\|^3_{\theta_0}\right] + 16\eta^2\|\mathrm{Hess}\,\mathrm{g}_\eta\|_\infty\,\mathbb{E}\left[\|Y_K\|^2_{\theta_0}\right] , \tag{S32}$$

where $\theta_0 = \mathrm{Exp}_{\theta^\star}(u_0)$, $\gamma:[0,1]\to\Theta$ is defined by $\gamma(t)=\mathrm{Exp}_{\theta_0}(t\eta H_{\theta_0}(X_1))$, H_K, Y_K and A_{θ_0} are defined in (S13). Using (S29) and Proposition S17, we have $\mathrm{Hess}\,\mathrm{g}_\eta(u) = \eta^{-1}\mathsf{C}^2(\mathrm{g},\eta^{-1/2}u_0,\eta)$ and $\nabla\mathrm{Hess}\,\mathrm{g}_\eta(u) = \eta^{-3/2}\mathsf{C}^3(\mathrm{g},\eta^{-1/2}u_0,\eta)$, where C^2 and C^3 are defined in (S24) and (S25) respectively. This gives

$$\|\nabla\mathrm{Hess}\,\mathrm{g}_\eta\|_{\gamma,\infty} = \eta^{-3/2}\|\mathsf{C}^3(\mathrm{g},\eta)\|_\gamma \quad \text{and} \quad \|\mathrm{Hess}\,\mathrm{g}_\eta\|_\infty = \eta^{-1}\|\mathsf{C}^2(\mathrm{g},\eta)\| , \tag{S33}$$

where $\|\mathsf{C}^2(\mathrm{g},\eta)\|$ and $\|\mathsf{C}^3(\mathrm{g},\eta)\|_\gamma$ are defined in (S28). Setting $u_0=\eta^{1/2}\overline{u}_0$ in (S31), we get

$$\begin{aligned} R_\eta \mathrm{g}_\eta(\eta^{1/2}\overline{u}_0) &= \mathrm{g}(\overline{u}_0) + \eta^{1/2}\sum_{i=1}^d \partial_i \mathrm{g}(\overline{u}_0)h^i\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)\right) \\ &+ (\eta/2)\sum_{i,j=1}^d \left\{\partial^2_{ij}\mathrm{g}\left(\overline{u}_0\right) - \eta^{1/2}\sum_{k=1}^d \Gamma^k_{ij}(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0))\partial_k \mathrm{g}\left(\overline{u}_0\right)\right\}\left[\Sigma_{ij}+h^ih^j\right]\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)\right) \\ &+ \eta^2\tilde{\mathscr{R}}_{\mathrm{g}_\eta,\eta}(\eta^{1/2}\overline{u}_0) . \end{aligned} \tag{S34}$$

Therefore, letting $\overline{\mathscr{R}}_{\mathrm{g},\eta}(\overline{u}_0) = \eta\tilde{\mathscr{R}}_{\mathrm{g}_\eta,\eta}(\eta^{1/2}\overline{u}_0)$, and combining Lemma S7-(S22), (S32), (S33) and (S34) gives the desired result. □

Lemma S9. *Assume* ***A*** *1-(i)-(ii) and* ***H*** *5. Consider normal coordinates $(u^i)_{i\in\{1,\dots,d\}}$ centered at $\theta^\star$ with respect to the orthonormal basis $(\mathbf{e}_i)_{i\in\{1,\dots,d\}}$ of $\mathrm{T}_{\theta^\star}\Theta$. Then h can be expressed in this chart as, for any $\eta>0$, $\overline{u}\in\mathrm{T}_{\theta^\star}\Theta$,*

$$h\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u})\right) = \sum_{i=1}^d\left\{\eta^{1/2}\sum_{k=1}^d \mathbf{A}^i_k\overline{u}^k + \mathscr{R}^i_h\left(\eta^{1/2}\overline{u}\right)\right\}\partial u_i , \tag{S35}$$

where $\mathbf{A}$ *is defined in* ***H*** *5, $\overline{u}^k$ are the components of $\overline{u}$ in $(\mathbf{e}_i)_{i\in\{1,\dots,d\}}$ and for any $i\in\{1,\dots,d\}$, $\lim_{u\to0}\{|\mathscr{R}^i_h(u)|/\|u\|_{\theta^\star}\}=0$.*

Proof. Since Θ is a Hadamard manifold, these normal coordinates are defined throughout Θ. Thus, for any $\theta\in\Theta$, it is possible to write,

$$h(\theta) = \sum_{j=1}^d h^j(\theta)\partial u_j(\theta) . \tag{S36}$$

Recall the definition of the metric coefficients in the coordinates $(u^i)_{i\in\{1,\dots,d\}}$ at $\theta\in\Theta$, for any $i,j\in\{1,\dots,d\}$,

$$\mathfrak{g}_{ij}(\theta) = \langle\partial u_i(\theta),\partial u_j(\theta)\rangle_\theta . \tag{S37}$$

Then, taking the scalar product of (S36) with each ∂u_i, we have for any $i\in\{1,\dots,d\}$,

$$\sum_{j=1}^d \mathfrak{g}_{ij}(\theta)\,h^j(\theta) = \langle h(\theta),\partial u_i(\theta)\rangle_\theta . \tag{S38}$$

From the Taylor expansion formula for vector fields given by Theorem S16 for the geodesic $\gamma:[0,1]\to\Theta$ given by $\gamma(0)=\theta^\star$ and $\dot\gamma(0)=\mathrm{Exp}^{-1}_{\theta^\star}(\theta)$, it follows that,

$$\partial u_i(\theta) = \mathrm{T}^\gamma_{01}\left[\mathbf{e}_i + \nabla(\partial u_i)_{\theta^\star}\left(\mathrm{Exp}^{-1}_{\theta^\star}(\theta)\right)\right] + \mathscr{R}_{\partial u_i}(\theta) , \tag{S39}$$

where the remainder is given by

$$\mathscr{R}_{\partial u_i}(\theta) = \int_0^1 (1-t)\mathrm{T}^\gamma_{t1}\nabla^2(\partial u_i)_{\gamma(t)}(\dot\gamma(t),\dot\gamma(t))\mathrm{d}t .$$

Let $\|\nabla^2 \partial u_i\|_{\infty,\gamma} = \sup\{|\nabla^2(\partial u_i)_{\gamma(t)}(v,v)| : t \in [0,1], v \in \mathrm{U}_{\gamma(t)}\Theta\}$ which is finite as $\gamma[0,1]$ is compact. Then using that for any $t \in [0,1]$, $\|\dot{\gamma}(t)\|_{\gamma(t)} = \rho_\Theta(\theta^\star,\theta)$ by [1, Corollary 5.6] and that geodesics are length-minimizing curves by **A**1-(i); and that the parallel transport map is an isometry [1, p.108], we have

$$|\mathscr{R}_{\partial u_i}(\theta)| \leq (1/2)\|\nabla^2 \partial u_i\|_{\infty,\gamma}\rho^2_\Theta(\theta^\star,\theta) .$$

This proves that $\lim_{\theta\to\theta^\star} |\mathscr{R}_{\partial u_i}(\theta)/\rho_\Theta(\theta^\star,\theta)| = 0$. By the definition of normal coordinates centered at $\theta^\star$, for any $i,j \in \{1,\ldots,d\}$, $\nabla_{\partial u_j}\partial u_i = \sum_{k=1}^d \Gamma^k_{ji}\partial u_k$ and $(\Gamma^k_{ji})_{i,j,k\in\{1,\ldots,d\}}$ vanishes at $\theta^\star$ [1, Proposition 5.24] so (S39) becomes

$$\partial u_i(\theta) = \mathrm{T}^\gamma_{01}(\mathbf{e}_i) + \mathscr{R}_{\partial u_i}(\theta) . \quad \text{(S40)}$$

Taking the scalar product of (12) and (S40), it follows that

$$\langle h(\theta), \partial u_i(\theta)\rangle_\theta = \langle \mathbf{A}\mathrm{Exp}^{-1}_{\theta^\star}(\theta), \mathbf{e}_i\rangle_{\theta^\star} + \tilde{\mathscr{R}}^i_h(\theta) , \quad \text{(S41)}$$

since parallel transport preserves scalar products, where $\lim_{\theta\to\theta^\star}\{|\tilde{\mathscr{R}}^i_h(\theta)|/\rho_\Theta(\theta^\star,\theta)\} = 0$. On the other hand, from (S37) and (S40), since the $(\mathbf{e}_i)_{i\in\{1,\ldots,d\}}$ are orthonormal,

$$\mathfrak{g}_{ij}(\theta) = \delta_{ij} + \mathscr{R}^{ij}_{\mathfrak{g}}(\theta) , \quad \text{(S42)}$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise and $\lim_{\theta\to\theta^\star}\{|\mathscr{R}^{ij}_{\mathfrak{g}}(\theta)|/\rho_\Theta(\theta^\star,\theta)\} = 0$. Plugging (S41) and (S42) in (S38), we obtain

$$h^i(\theta) = \sum_{j=1}^d \mathbf{A}^i_j u^j(\theta) + \mathscr{R}^i_h(\theta) , \quad \text{(S43)}$$

where $\lim_{\theta\to\theta^\star}\left|\mathscr{R}^i_h(\theta)\right| = 0$. Finally, (S35) is obtained from (S36)-(S43), by setting $\theta = \mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u})$, for $\overline{u} \in \mathrm{T}_{\theta^\star}\Theta$, and noting that

$$u^j(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u})) = \langle \mathrm{Exp}^{-1}_{\theta^\star}(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u})), \mathbf{e}_j\rangle_{\theta^\star} = \eta^{1/2}\overline{u}^j ,$$
$$\rho_\Theta\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}), \theta^\star\right) = \eta^{1/2}\|\overline{u}\|_{\theta^\star} ,$$

which follow from [1, Corollary 5.6] and the definition of the coordinates $(u^i)_{i\in\{1,\ldots,d\}}$. ☐

Lemma S10. *Assume **A** 1-(i)-(ii), **MD** 1, **MD** 2, **MD** 3, **MD** 4, **H** 1, **H** 2, **H** 5 and **H** 6 hold. Let* $\overline{\eta} = [2C_2L(1+\sigma_1^2)]^{-1} \wedge (4C_3)^{-1}$. *Then the family of distributions* $(\overline{\nu}^\eta)_{\eta\in(0,\overline{\eta}]}$, *defined by* (11), *is tight.*

Proof. For any $\eta \in (0,\overline{\eta}]$, the conditions of Lemma S7 hold, thus the Markov chain $(\overline{U}_n)_{n\in\mathbb{N}}$ is ergodic and its invariant distribution $\overline{\nu}^\eta$ is given by (11). For any $r \geq 0$, let $\overline{\mathbb{B}}_r = \{u \in \mathrm{T}_{\theta^\star}\Theta : \|u\|_{\theta^\star} \leq r\}$ be the tangent closed ball at $\theta^\star$ of center 0 and radius r. Then, by (S23) and [1, Corollary 6.13], for any $r > 0$ and $\eta \in (0,\overline{\eta}]$, we have

$$\overline{\nu}^\eta\left(\mathrm{T}_{\theta^\star}\Theta \setminus \overline{\mathbb{B}}_r\right) = \nu^\eta\left(\mathrm{T}_{\theta^\star}\Theta \setminus \overline{\mathbb{B}}_{\eta^{1/2}r}\right) = \mu^\eta\left(\Theta \setminus \overline{\mathrm{B}}(\theta^\star, \eta^{1/2}r)\right) . \quad \text{(S44)}$$

However, by **H**6,

$$\begin{aligned}\mu^\eta\left(\Theta \setminus \overline{\mathrm{B}}(\theta^\star,\eta^{1/2}r)\right) &\leq \phi^{-1}(\eta^{1/2}r)\int_{\Theta\setminus\{\theta^\star\}} \phi(\rho_\Theta(\theta^\star,\theta))\mathrm{d}\mu^\eta(\theta) \\ &\leq \phi^{-1}(\eta^{1/2}r)\int_{\Theta\setminus\{\theta^\star\}} V(\theta)\mathrm{d}\mu^\eta(\theta) . \end{aligned} \quad \text{(S45)}$$

Now, using **H**6 and Lemma S5 taking $\mathsf{K}^\star = \{\theta^\star\}$, we have,

$$\int_{\Theta\setminus\{\theta^\star\}} V(\theta)\mathrm{d}\mu^\eta(\theta) \leq 2\eta L\left\{\sigma_0^2 + C_1(1+\sigma_1^2)\right\}/\lambda .$$

Combining this result and (S45) in (S44) implies that for any $r > 0$,

$$\begin{aligned}\overline{\nu}^\eta\left(\mathrm{T}_{\theta^\star}\Theta \setminus \overline{\mathbb{B}}_r\right) &\leq 2\eta L\left\{\sigma_0^2 + C_1(1+\sigma_1^2)\right\}/[\lambda\phi(\eta^{1/2}r)] \\ &\leq \sup_{\eta\leq\overline{\eta}}\{\eta/\phi(\eta^{1/2}r)\}(2L/\lambda)\left\{\sigma_0^2 + C_1(1+\sigma_1^2)\right\} ,\end{aligned}$$

where $\lim_{r\to+\infty}\{\sup_{\eta\leq\overline{\eta}}\eta/\phi(\eta^{1/2}r)\} = 0$ using **H**6. Therefore, for any $\varepsilon > 0$, there exists $r > 0$ such that for any $\eta \in (0,\overline{\eta}]$, $\overline{\nu}^\eta(\mathrm{T}_{\theta^\star}\Theta \setminus \overline{\mathbb{B}}_r) \leq \varepsilon$. This concludes the proof that $(\overline{\nu}^\eta)_{\eta\in(0,\overline{\eta}]}$ is tight. ☐

Proof of Theorem 7. Consider normal coordinates $(u^i)_{i\in\{1,\ldots,d\}}$ centered at $\theta^\star$ with respect to the orthonormal basis $(\mathbf{e}_i)_{i\in\{1,\ldots,d\}}$ of $\mathrm{T}_{\theta^\star}\Theta$. Define for any $i,j \in \{1,\ldots,d\}$, $h^i : \Theta \to \mathbb{R}$, $\Sigma_{ij} : \Theta \to \mathbb{R}$ by $h^i = \mathrm{d}u^i(h)$ and $\Sigma_{ij} = [\mathrm{d}u^i \otimes \mathrm{d}u^j]\{\Sigma\}$. Let $\mathrm{g} : \mathrm{T}_{\theta^\star}\Theta \to \mathbb{R}$ be a smooth function with compact support. Applying Lemma S8 to g gives (S26). Using **MD**3, Σ is continuous, which implies that for any $\overline{u}_0 \in \mathrm{T}_{\theta^\star}\Theta$,

$$\Sigma\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)\right) = \sum_{i,j=1}^{d}\left\{\Sigma_\star^{ij} + \mathscr{R}_\Sigma^{ij}\left(\eta^{1/2}\overline{u}_0\right)\right\}\partial u_i \otimes \partial u_j \;, \tag{S46}$$

where for any $i,j \in \{1,\ldots,d\}$, $\Sigma_\star^{ij} = \Sigma_{ij}(\theta^\star)$, $\mathscr{R}_\Sigma^{ij}$ is continuous over $\mathrm{T}_{\theta^\star}\Theta$ and $\mathscr{R}_\Sigma^{ij}(0) = 0$. Using Lemma S9, replacing Σ_{ij} and h^i in (S26) with (S35) and (S46) gives for any $\overline{u}_0 \in \mathrm{T}_{\theta^\star}\Theta$,

$$\begin{aligned}\overline{R}_\eta \mathrm{g}(\overline{u}_0) = \mathrm{g}(\overline{u}_0) + \eta\sum_{i=1}^d \partial_i \mathrm{g}(\overline{u}_0)\sum_{k=1}^d \mathbf{A}_k^i \overline{u}_0^k + (\eta/2)\sum_{i,j=1}^d \partial_{ij}^2 \mathrm{g}(\overline{u}_0)\Sigma_\star^{ij} + \eta\mathscr{R}_{\mathrm{g},\eta,\Sigma,h}(\overline{u}_0) \\ + (\eta/6)\overline{\mathscr{R}}_{\mathrm{g},\eta}(\overline{u}_0)\;,\end{aligned} \tag{S47}$$

where $\overline{u}_0^k$ are the components of $\overline{u}_0$ in $(\mathbf{e}_i)_{i\in\{1,\ldots,d\}}$,

$$\begin{aligned}\mathscr{R}_{\mathrm{g},\eta,\Sigma,h}(\overline{u}_0) &= \eta^{-1/2}\sum_{i=1}^d \mathscr{R}_h^i\left(\eta^{1/2}\overline{u}_0\right)\partial_i \mathrm{g}(\overline{u}_0) \\ &+ (1/2)\sum_{i,j=1}^d\left\{\partial_{ij}^2\mathrm{g}(\overline{u}_0) - \eta^{1/2}\sum_{k=1}^d \Gamma_{ij}^k\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)\right)\partial_k \mathrm{g}(\overline{u}_0)\right\}\left[\mathscr{R}_\Sigma^{ij}\left(\eta^{1/2}\overline{u}_0\right)\right] \\ &+ (1/2)\sum_{i,j=1}^d\left\{\partial_{ij}^2\mathrm{g}(\overline{u}_0) - \eta^{1/2}\sum_{k=1}^d \Gamma_{ij}^k\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)\right)\partial_k \mathrm{g}(\overline{u}_0)\right\}\left[h^i h^j\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)\right)\right] \\ &- (\eta^{1/2}/2)\sum_{i,j,k=1}^d \Gamma_{ij}^k\left(\mathrm{Exp}_{\theta^\star}(\eta^{1/2}\overline{u}_0)\right)\partial_k \mathrm{g}(\overline{u}_0)\Sigma_\star^{ij}\;.\end{aligned}$$

By Lemma S10, $(\overline{\nu}^\eta)_{\eta\in(0,\overline{\eta}]}$ is tight and therefore relatively compact. Therefore, it is enough that for any limit point $\overline{\nu}^\star$, $\overline{\nu}^\star = \mathrm{N}(0,\mathbf{V})$ where $\mathbf{V} \in \mathbb{R}^{d\times d}$ is the solution of the Lyapunov equation $\mathbf{AV} + \mathbf{VA}^\top = \Sigma(\theta^\star)$. Let $(\eta_n)_{n\in\mathbb{N}^*}$ be a sequence with values in $(0,\overline{\eta}]$, such that $\lim_{n\to+\infty}\eta_n = 0$, and $(\overline{\nu}^{\eta_n})_{n\in\mathbb{N}^*}$ weakly converges to $\overline{\nu}^\star$.

First by (S47), we have

$$\begin{aligned}&\int_{\mathrm{T}_{\theta^\star}\Theta}\overline{\nu}^{\eta_n}(\mathrm{d}\overline{u}_0)\int_{\mathrm{T}_{\theta^\star}\Theta}\overline{R}_{\eta_n}(\overline{u}_0,\mathrm{d}\overline{u}_1)\mathrm{g}(\overline{u}_1) \\ &= \int_{\mathrm{T}_{\theta^\star}\Theta}\overline{\nu}^{\eta_n}(\mathrm{d}\overline{u}_0)\mathrm{g}(\overline{u}_0) + \eta_n\int_{\mathrm{T}_{\theta^\star}\Theta}\overline{\nu}^{\eta_n}(\mathrm{d}\overline{u}_0)\sum_{i=1}^d\partial_i\mathrm{g}(\overline{u}_0)\sum_{k=1}^d\mathbf{A}_k^i\overline{u}_0^k \\ &\quad + (\eta_n/2)\int_{\mathrm{T}_{\theta^\star}\Theta}\overline{\nu}^{\eta_n}(\mathrm{d}\overline{u}_0)\sum_{i,j=1}^d\partial_{ij}^2\mathrm{g}(\overline{u}_0)\Sigma_\star^{ij} + \eta_n\int_{\mathrm{T}_{\theta^\star}\Theta}\overline{\nu}^{\eta_n}(\mathrm{d}\overline{u}_0)\mathscr{R}_{\mathrm{g},\eta_n,\Sigma,h}(\overline{u}_0) \\ &\quad + (\eta_n/6)\int_{\mathrm{T}_{\theta^\star}\Theta}\overline{\nu}^{\eta_n}(\mathrm{d}\overline{u}_0)\overline{\mathscr{R}}_{\mathrm{g},\eta_n}(\overline{u}_0)\;.\end{aligned}$$

Therefore using that $\overline{\nu}^{\eta_n}$ is stationary with respect to $\overline{R}_{\eta_n}$, we obtain that

$$\begin{aligned}\limsup_{n\to+\infty}&\left|\int_{\mathrm{T}_{\theta^\star}\Theta}\overline{\nu}^{\eta_n}(\mathrm{d}\overline{u}_0)\left\{\sum_{i=1}^d\partial_i\mathrm{g}(\overline{u}_0)\sum_{k=1}^d\mathbf{A}_k^i\overline{u}_0^k + \sum_{i,j=1}^d\partial_{ij}^2\mathrm{g}(\overline{u}_0)\Sigma_\star^{ij}\right\}\right| \\ &\le \limsup_{n\to+\infty}\left|\int_{\mathrm{T}_{\theta^\star}\Theta}\overline{\nu}^{\eta_n}(\mathrm{d}\overline{u}_0)\mathscr{R}_{\mathrm{g},\eta_n,\Sigma,h}(\overline{u}_0)\right| + \left|\int_{\mathrm{T}_{\theta^\star}\Theta}\overline{\nu}^{\eta_n}(\mathrm{d}\overline{u}_0)\overline{\mathscr{R}}_{\mathrm{g},\eta_n}(\overline{u}_0)\right|\;.\end{aligned} \tag{S48}$$

Consider a sequence of independent random variables $(\mathrm{Y}_n)_{n\in\mathbb{N}}$ such that for any $n \in \mathbb{N}$, the law of Y_n is $\overline{\nu}^{\eta_n}$. By Slutsky's theorem, since $(\mathrm{Y}_n)_{n\in\mathbb{N}}$ converges in distribution and $\lim_{n\to+\infty} \eta_n = 0$, we obtain that $\eta_n^{1/2}\mathrm{Y}_n$ converges in distribution towards 0. Moreover, using the continuous mapping theorem, we have

$$\limsup_{n\to+\infty} |\mathbb{E}[\mathscr{R}_{\mathrm{g},\eta_n,\Sigma,h}(\mathrm{Y}_n)]| = 0 . \tag{S49}$$

Similarly, we use (S27) to obtain, for any $n \in \mathbb{N}$ and $K > 0$,

$$\begin{aligned}\left|\overline{\mathscr{R}}_{\mathrm{g},\eta_n}(\mathrm{Y}_n)\right| \le\ & 8\eta_n^{1/2}\mathbb{1}_{\mathsf{K}_K}(\theta_n)\mathbb{E}\left[\|\mathsf{C}^3(\mathrm{g},\eta_n)\|_\gamma \mathbb{1}_{\mathsf{A}^\complement_{\theta_n}} \|H_K\|^3_{\theta_n}\middle|\theta_n\right] \\ & + 16\|\mathsf{C}^2(\mathrm{g},\eta_n)\|\mathbb{E}\left[\|Y_K\|^2_{\theta_n}\middle|\theta_n\right] ,\end{aligned}$$

where for any $n \in \mathbb{N}$, $\theta_n = \mathrm{Exp}_{\theta^\star}(\eta_n^{1/2}\mathrm{Y}_n)$ are independent random variables and by (S23), the distribution of θ_n is μ^{η_n}. Thus we obtain for any $K \ge 0$, using $\mathbb{1}_{\mathsf{K}_K}(\theta_n)\,\|H_K\|_{\theta_n}$ is almost surely bounded by $4[K^3 + \sup_{\theta\in\mathsf{K}_K}\|h(\theta)\|^3_\theta]$, Markov's inequality and **MD**4,

$$\begin{aligned}\limsup_{n\to+\infty}\left|\mathbb{E}[\overline{\mathscr{R}}_{\mathrm{g},\eta_n}(\mathrm{Y}_n)]\right| &\le \limsup_{n\to+\infty} 16\|\mathsf{C}^2(\mathrm{g},\eta_n)\|\mathbb{E}[\|e_{\theta_n}(X_1)\|^2_{\theta_n}\{1-\chi_K(\|e_{\theta_n}(X_1)\|_{\theta_n})\}] \\ &\le 16\|\mathsf{C}^2(\mathrm{g},0)\|K^{-\varepsilon}\{\tilde{\sigma}_0^2 + \tilde{\sigma}_1^2\mathbb{E}[V(\theta^\star)]\} ,\end{aligned} \tag{S50}$$

using that $(\theta_n)_{n\in\mathbb{N}}$ converges in distribution to $\theta^\star$. For any smooth function with compact support $\mathrm{g} : \mathrm{T}_{\theta^\star}\Theta \to \mathbb{R}$, combining (S48)-(S49)-(S50), taking $K \to +\infty$ and using the weak convergence of $(\overline{\nu}^{\eta_n})_{n\in\mathbb{N}}$ to $\overline{\nu}^\star$ when $n \to +\infty$ shows that

$$\int_{\mathrm{T}_{\theta^\star}\Theta} \overline{\nu}^\star(\mathrm{d}\overline{u}_0)\left\{\sum_{i=1}^d \partial_i \mathrm{g}(\overline{u}_0)\sum_{k=1}^d \mathbf{A}^i_k \overline{u}^k_0 + \sum_{i,j=1}^d \partial^2_{ij}\mathrm{g}(\overline{u}_0)\Sigma^{ij}_\star\right\} = 0 . \tag{S51}$$

Finally, by [6, Theorem 2.2.1], there exists a unique matrix $\mathbf{V} \in \mathbb{R}^{d\times d}$ solution to the Lyapunov equation $\mathbf{AV} + \mathbf{VA}^\top = \Sigma(\theta^\star)$. By [7, Theorem 10.1], $\mathrm{N}(0, \mathbf{V})$ is the unique probability distribution on $\mathrm{T}_{\theta^\star}\Theta$ satisfying (S51). This concludes the proof. □

S5 Proofs for Section 4

S5.1 Proof of Lemma 8

Recall that f is λ_f-strongly geodesically convex, if and only if for any $\theta_1, \theta_2 \in \Theta$,

$$f(\theta_2) \ge f(\theta_1) + \left\langle \mathrm{Exp}^{-1}_{\theta_1}(\theta_2), \mathrm{grad}\, f(\theta_1)\right\rangle_{\theta_1} + \lambda_f \rho^2_\Theta(\theta_1, \theta_2) . \tag{S52}$$

Put $\theta_1 = \theta^\star$ and $\theta_2 = \theta$. Since $\theta^\star$ is a stationary point of f, so $\mathrm{grad}\, f(\theta^\star) = 0$, it follows from (S52) that

$$f(\theta) - f(\theta^\star) \ge \lambda_f \rho^2_\Theta(\theta^\star, \theta) ,$$

which is the second identity in (13). To obtain the first identity, put $\theta_1 = \theta$ and $\theta_2 = \theta^\star$, in (S52), so

$$f(\theta^\star) - f(\theta) \ge \left\langle \mathrm{Exp}^{-1}_\theta(\theta^\star), \mathrm{grad}\, f(\theta)\right\rangle_\theta + \lambda_f \rho^2_\Theta(\theta^\star, \theta) . \tag{S53}$$

Since $f(\theta^\star) \le f(\theta)$, this implies

$$-\left\langle \mathrm{Exp}^{-1}_\theta(\theta^\star), \mathrm{grad}\, f(\theta)\right\rangle_\theta \ge \lambda_f \rho^2_\Theta(\theta^\star, \theta) = \lambda_f \left\|\mathrm{Exp}^{-1}_\theta(\theta^\star)\right\|^2_\theta .$$

Or, after using the Cauchy-Schwarz inequality,

$$\|\mathrm{grad}\, f(\theta)\|_\theta \ge \lambda_f \left\|\mathrm{Exp}^{-1}_\theta(\theta^\star)\right\|_\theta . \tag{S54}$$

Finally, using once more the Cauchy-Schwarz inequality, and (S53) and (S54),

$$f(\theta) - f(\theta^\star) \le -\left\langle \mathrm{Exp}^{-1}_\theta(\theta^\star), \mathrm{grad}\, f(\theta)\right\rangle_\theta \le (1/\lambda_f)\,\|\mathrm{grad}\, f(\theta)\|^2_\theta ,$$

which is equivalent to the first identity in (13).

S5.2 Proof of Lemma 10

Without loss of generality, we assume that $f(\theta^\star) = 0$. First, we show that for any $\theta \in \Theta$,

$$f(\theta) \leq M_f \rho_\Theta^2(\theta^\star, \theta) . \tag{S55}$$

Let $\theta \in \Theta$ and $\gamma : [0,1] \to \Theta$ the unique geodesic such that $\gamma(0) = \theta^\star$ and $\gamma(1) = \theta$. Then since f is continuously differentiable using [1, Proposition 4.15-(ii) and Theorem 4.24-(iii)], we get that $f(\theta) = \int_0^1 \langle \mathrm{grad}\, f(\gamma(t)), \dot\gamma(t)\rangle_{\gamma(t)}\, \mathrm{d}t$. Therefore, using the Cauchy-Schwarz inequality and for any $t \in [0,1]$, $\|\dot\gamma(t)\|_{\gamma(t)} = \rho_\Theta(\theta^\star, \theta)$ we obtain that $|f(\theta)| \leq \rho_\Theta(\theta^\star,\theta) \|\mathrm{grad}\, f(\gamma(t))\|_{\gamma(t)}$ which shows that (S55) holds by assumption.

We now proceed with the proof of the main statement. Since f is twice continuously differentiable, $\tilde f$ has this same property. In addition, for any $\theta \in \Theta$,

$$\mathrm{grad}\, \tilde f(\theta) = \mathrm{grad}\, f(\theta)/[2(f(\theta)+1)^{1/2}] . \tag{S56}$$

Therefore, using the assumption that for any $\theta \in \Theta$, $\|\mathrm{grad}\, f(\theta)\|_\theta^2 \leq M_f \rho_\Theta^2(\theta^\star, \theta)$ and the second inequality of Lemma 8, we get that

$$\begin{aligned}\|\mathrm{grad}\, \tilde f(\theta)\|_\theta = \|\mathrm{grad}\, f(\theta)\|_\theta /[2(f(\theta)+1)^{1/2}] \leq M_f^{1/2} \rho_\Theta(\theta^\star,\theta)/[2(\lambda_f \rho_\Theta^2(\theta^\star,\theta)+1)^{1/2}] \\ \leq C_f^{1/2}[1 \wedge \rho_\Theta(\theta^\star,\theta)] ,\end{aligned}$$

with $C_f^{1/2} \leftarrow (M_f^{1/2}/2)[1 \wedge \lambda_f^{-1/2}]$.

It remains to show that for any $\theta \in \Theta$, $-\langle \mathrm{Exp}_\theta^{-1}(\theta^\star), \mathrm{grad}\, \tilde f(\theta)\rangle_\theta \geq \tilde\lambda_f V_1(\theta)$, where V_1 is defined by (9) with $\delta = 1$ and $\tilde\lambda_f \leftarrow \lambda_f^{1/2}/2$. Using (S56) again, **F**2 and (S55), we obtain that for any $\theta \in \Theta$,

$$\begin{aligned}-\left\langle \mathrm{Exp}_\theta^{-1}(\theta^\star), \mathrm{grad}\, \tilde f(\theta)\right\rangle_\theta &= -\left\langle \mathrm{Exp}_\theta^{-1}(\theta^\star), \mathrm{grad}\, f(\theta)\right\rangle_\theta /[2(f(\theta)+1)^{1/2}] \\ &\geq \lambda_f \rho_\Theta^2(\theta^\star,\theta)/[2(f(\theta)+1)^{1/2}] \geq \lambda_f \rho_\Theta^2(\theta^\star,\theta)/[2(M_f\rho_\Theta^2(\theta^\star,\theta)+1)^{1/2}] .\end{aligned}$$

Using that for any $\theta \in \Theta$, $V_1(\theta) = \{\rho_\Theta^2(\theta^\star,\theta)+1\}^{1/2} - 1 \leq \rho_\Theta(\theta^\star,\theta)$, we get that

$$-\left\langle \mathrm{Exp}_\theta^{-1}(\theta^\star), \mathrm{grad}\, \tilde f(\theta)\right\rangle_\theta \geq \lambda_f V_1(\theta)\rho_\Theta(\theta^\star,\theta)/[2(M_f\rho_\Theta^2(\theta^\star,\theta)+1)^{1/2}] \geq \lambda_f V_1(\theta)/(2M_f^{1/2}) .$$

S5.3 Proof of Proposition 11

The proof consists in an application of Theorem 1-(b). First, by Proposition 5, V_1 defined by (9) with $\delta = 1$, satisfies **H**1. In addition, by [2, Lemma 16], V_1 is continuously differentiable with gradient given for any $\theta \in \Theta$ by

$$\mathrm{grad}\, V_1(\theta) = -\mathrm{Exp}_\theta^{-1}(\theta^\star)/\{1+\rho_\Theta^2(\theta^\star,\theta)\}^{1/2} .$$

Therefore, for any $\theta \in \Theta$, by **F**3 we get

$$\begin{aligned}\langle \mathrm{grad}\, V_1(\theta), \mathrm{grad}\, f(\theta)\rangle_\theta = -\left\langle \mathrm{Exp}_\theta^{-1}(\theta^\star), \mathrm{grad}\, f(\theta)\right\rangle_\theta /\{1+\rho_\Theta^2(\theta^\star,\theta)\}^{1/2} \\ \geq \tilde\lambda_f V_1(\theta)/\{1+\rho_\Theta^2(\theta^\star,\theta)\}^{1/2} .\end{aligned} \tag{S57}$$

In addition, $t^2 \wedge 1 - ab\{(t^2+1)^{1/2} - 1\}/(1+t^2)^{1/2} \leq 0$ for any $t \geq 0$, $b > 0$ and $a = 4b^{-1}$ using that $(t^2+1)^{1/2} - 1 \geq t^2/[2(1+t^2)^{1/2}]$. As a result, using **F**3 for any $t \geq 0$, $b > 0$ and $a = 4b^{-1}$, it follows that **H**2 is satisfied with $C_1 \leftarrow 0, C_2 \leftarrow 4C_f/\tilde\lambda_f$ for $h = -\mathrm{grad}\, f$ and $V \leftarrow V_1$. Therefore, we obtain using Theorem 1-(b) that for any $\eta \in (0, \eta]$,

$$n^{-1}\sum_{k=0}^{n-1} \mathbb{E}\left[\langle \mathrm{grad}\, V_1(\theta_k), \mathrm{grad}\, f(\theta_k)\rangle_{\theta_k}\right] \leq 2V_1(\theta_0)/(n\eta) + 2\eta(1+\kappa)\sigma_0^2 ,$$

where $\overline{\eta} = [(8C_f/\tilde{\lambda}_f)(1+\kappa)(1+\sigma_1^2)]^{-1}$. Using (S57), we have

$$(\tilde{\lambda}_f/n)\sum_{k=0}^{n-1}\mathbb{E}\left[V_1(\theta_k)/\{1+\rho_\Theta^2(\theta^\star,\theta_k)\}^{1/2}\right] \leq 2V_1(\theta_0)/(n\eta) + 2\eta(1+\kappa)\sigma_0^2 \ ,$$

which concludes the proof since $(t^2+1)^{1/2}-1 \geq t^2/[2(1+t^2)^{1/2}]$ for any $t \geq 0$ implying $V_1(\theta)/\{1+\rho_\Theta^2(\theta^\star,\theta)\}^{1/2} \geq D_\Theta^2(\theta^\star,\theta)/2$ for any $\theta \in \Theta$.

S5.4 Proof of Proposition 12

Define $\mathsf{X} = \{\overline{\theta}_i \,:\, i \in \{1,\ldots,M_\pi\}\}$ and recall that $\mathrm{D} = \sup\{\rho_\Theta(\theta_0,\overline{\theta}) \,:\, \overline{\theta} \in \mathsf{X}\}$. Set $\mathsf{S} = \overline{\mathrm{B}}(\theta_0,\mathrm{D})$. Note that the closed ball S, is compact by [4, Theorem 1.7.1], geodesically convex, and $\mathsf{X} \subset \mathsf{S}$, as well as $\theta_0 \in \mathsf{S}$. We consider in this section, for any $\theta \in \Theta$ and $x \in \mathsf{X}$, $H_\theta(x) = \mathrm{Exp}_\theta^{-1}(x)$.

First note that $\theta_n \in \mathsf{S}$, for all $n \in \mathbb{N}$ by a straightforward induction using that S is geodesically convex and $\theta_0 \in \mathsf{S}$. Indeed, $\theta_0 \in \mathsf{S}$, and, if $\theta_n \in \mathsf{S}$, then θ_{n+1} lies on the geodesic segment connecting θ_n and X_{n+1}, two points which belong to S, and therefore $\theta_{n+1} \in \mathsf{S}$. This means that the SGD scheme used here is equivalent to

$$\theta_{n+1} = \mathrm{proj}_{\mathsf{S}}\left(\mathrm{Exp}_{\theta_n}(\eta H_{\theta_n}(X_{n+1}))\right) \ .$$

Define H and V_2 as in Proposition 4. It is possible to show that $\mathsf{H} = \mathsf{S}$. Indeed, for $\theta \in \mathsf{S}$, and $x \in \mathsf{X}$, since $x \in \mathsf{S}$, and S is convex, the geodesic segment connecting θ to x is entirely contained in S. However, by definition, this geodesic segment is the set of points $\mathrm{Exp}_\theta(tH_\theta(x))$, where $t \in [0,1]$. Now, since $\eta \leq \overline{\eta} \leq 1$, Proposition 4 implies that V_2 verifies **H**1-(i)-(ii) where $L \leftarrow CL_\pi$, $L_\pi = (\mathrm{D}+1)(1+\kappa\coth(\kappa\mathrm{D}))$ and C is a universal constant.

The objective function f satisfies **F**2 with $\lambda_f = 1/2$ (that is, f is 1/2-strongly convex), since by [4, Theorem 5.6.1] $f_i(\theta) = \rho_\Theta^2(\theta,\overline{\theta}_i)/2$ is 1-strongly geodesically convex for any $i \in \{1,\ldots,M_\pi\}$. Thus, by (S52) for all $\theta \in \mathsf{S}$

$$\left\langle \mathrm{Exp}_\theta^{-1}(\theta^\star), \mathrm{grad}\, f(\theta)\right\rangle_\theta \leq -(1/2)\rho_\Theta^2(\theta^\star,\theta) \ . \tag{S58}$$

Now, for any $\theta \in \mathsf{S}$, $v \in \mathrm{T}_\theta\Theta$, using [4, Theorem 5.6.1], we have,

$$\begin{aligned}\|\mathrm{Hess}\, f_\theta(v,v)\|_\theta &\leq M_\pi^{-1}\sum_{i=1}^{M_\pi}\|(\mathrm{Hess}\, f_i)_\theta(v,v)\|_\theta \\ &\leq M_\pi^{-1}\sum_{i=1}^{M_\pi}\kappa\rho_\Theta(\theta,\overline{\theta}_i)\coth(\kappa\rho_\Theta(\theta,\overline{\theta}_i))\,\|v\|_\theta^2 \leq \tilde{L}_\pi\,\|v\|_\theta^2 \ ,\end{aligned}$$

where $\tilde{L}_\pi = 2\mathrm{D}\kappa\coth(2\kappa\mathrm{D})$, since $t \mapsto t\coth(t)$ is non-decreasing over $\mathbb{R}_+$. Therefore, by [2, Lemma 10], grad f is geodesically $\tilde{L}_\pi$-Lipschitz continuous on S.In particular, for any $\theta \in \mathsf{S}$,

$$\|\mathrm{grad}\, f(\theta)\|_\theta \leq \tilde{L}_\pi\rho_\Theta(\theta^\star,\theta) \ . \tag{S59}$$

By (S58) and (S59), it is straightforward that $V = V_2$ and $h = -\mathrm{grad}\, f$ satisfy **H**2, with $C_1 = 0$ and $C_2 = 2\tilde{L}_\pi^2 \leq 2^5 L_\pi^2$. In addition, by Proposition 4, (S58) implies V_2 verifies **H**3-($\emptyset$), with $\lambda = 1/2$.

Finally, **MD**1 holds with $\sigma_0^2 = \mathrm{D}^2$ and $\sigma_1^2 = 0$ since for any $\theta \in \mathsf{S}$ and $x \in \mathsf{X}$,

$$\|H_\theta(x)\|_\theta = \left\|\mathrm{Exp}_\theta^{-1}(x)\right\|_\theta \leq 2\mathrm{D} \ .$$

Therefore, we can apply Theorem 1-(c) which implies that for any $\eta \leq \overline{\eta}$,

$$\mathbb{E}[V_2(\theta_n)] \leq \{1-\eta/4\}^n\, V_2(\theta_0) + 4\eta\, L_\pi\mathrm{D}^2 \ .$$

To conclude, it only remains to note that $V_2(\theta_n) = \rho_\Theta^2(\theta^\star,\theta_n)$ and $V_2(\theta_0) = \rho_\Theta^2(\theta^\star,\theta_0)$, since $(\theta_n)_{n\in\mathbb{N}}$ and $\theta^\star$ belong to $\mathsf{H} = \mathsf{S}$.

S5.5 Proof of Theorem 13

We consider in this section the recursion

$$\theta_{n+1} = \mathrm{Exp}_{\theta_n}\left[\eta H_{\theta_n}(X_{n+1})\right] \tag{S60}$$
$$H_{\theta_n}(X_{n+1}) = \mathrm{Exp}_{\theta_n}^{-1}\left(X_{n+1}^{(1)}\right)\Big/\left(2\{\rho_\Theta^2(\theta_n, X_{n+1}^{(2)})/2+1\}^{1/2}\right) ,$$

where $X_{n+1} = (X_{n+1}^{(1)}, X_{n+1}^{(2)}$ and $(X_n^{(1)}, X_n^{(2)})_{n\in\mathbb{N}^*}$ is an i.i.d. sequence of pairs of independent random variables with distribution π. Denote by Q_η the Markov kernel corresponding to (S60).

We give first some additional intuition and motivation behind the scheme (S60). It can be interpreted as a stochastic optimization method to minimize

$$\tilde{f}_\pi = (f_\pi + 1)^{1/2} ,$$

in place of f_π. First note that f_π and $\tilde{f}_\pi$ have the same minimizer, but compared to f_π it may be shown that $\mathrm{grad}\,\tilde{f}_\pi$, given for any $\theta \in \Theta$ by

$$\mathrm{grad}\,\tilde{f}_\pi(\theta) = (1/2)\mathrm{grad}\,f_\pi(\theta)(f_\pi(\theta)+1)^{-1/2} ,$$

is geodesically Lipschitz. However, note that (S60) is not an unbiased stochastic optimization scheme for the function $\tilde{f}_\pi$ since

$$\mathbb{E}\left[H_{\theta_n}(X_{n+1})\right] = (1/2)\{\mathrm{grad}\,f_\pi(\theta_n)\}\mathbb{E}\left[\{\rho_\Theta^2(\theta_n, X_{n+1}^{(2)})/2+1\}^{-1/2}\right] .$$

The proof of Theorem 13 then consists in adapting the proof of Theorem 1 to deal with this additional difficulty taking for the Lyapunov function V, V_1 defined by (9) with $\delta = 1$. A general theory could be derived but we believe that this is out the scope of the present document and leave it for future work. We start by preliminary technical results which are needed to establish Theorem 13.

Lemma S11. *Assume* **A** *2 and* **MD** *5. Let $\theta_\pi^\star$ be the Riemannian barycenter of the probability measure π, i.e. $\theta_\pi^\star = \mathrm{argmin}_\Theta f_\pi$ where f_π is defined by (16). Then, for any $\theta \in \Theta$,*

$$-\int_\Theta \left\langle \mathrm{Exp}_\theta^{-1}(\theta_\pi^\star), \mathrm{Exp}_\theta^{-1}(\nu)\right\rangle_\theta \pi(\mathrm{d}\nu) \le -\rho_\Theta^2(\theta, \theta_\pi^\star)/2 .$$

Proof. Using **A**2 and [4, Theorem 5.6.1], we have that for any $\nu \in \Theta$, the operator norm of the Riemannian Hessian of $\theta \mapsto \rho_\Theta^2(\theta,\nu)/2$ is lower bounded by 1. Therefore, by [8, Theorem 11.19], $\theta \mapsto \rho_\Theta^2(\theta,\nu)/2$ is 1/2-strongly convex. Applying this to θ and $\theta_\pi^\star \in \Theta$, we have for any $\nu \in \Theta$,

$$\rho_\Theta^2(\theta_\pi^\star,\nu)/2 - \rho_\Theta^2(\theta,\nu)/2 \ge -\left\langle \mathrm{Exp}_\theta^{-1}(\theta_\pi^\star), \mathrm{Exp}_\theta^{-1}(\nu)\right\rangle_\theta + \rho_\Theta^2(\theta,\theta_\pi^\star)/2 .$$

Using **MD**5, we can integrate this inequality w.r.t. π, bringing

$$f_\pi(\theta_\pi^\star) - f_\pi(\theta) \ge -\int_\Theta \left\langle \mathrm{Exp}_\theta^{-1}(\theta_\pi^\star), \mathrm{Exp}_\theta^{-1}(\nu)\right\rangle_\theta \pi(\mathrm{d}\nu) + \rho_\Theta^2(\theta,\theta_\pi^\star)/2 .$$

Since by definition of $\theta_\pi^\star$, $0 \ge f_\pi(\theta_\pi^\star) - f_\pi(\theta)$, this completes the proof. □

Lemma S12. *Assume* **A** *2 and* **MD** *5. Let $\theta_\pi^\star$ be the Riemannian barycenter of the probability measure π, i.e. $\theta_\pi^\star = \mathrm{argmin}_\Theta f_\pi$ where f_π is defined by (16). Then, for any $\theta \in \Theta$,*

$$\int_\Theta \{\rho_\Theta^2(\theta,\nu)/2+1\}^{-1/2}\pi(\mathrm{d}\nu) \ge \{\rho_\Theta^2(\theta,\theta_\pi^\star) + 2f_\pi(\theta_\pi^\star)+1\}^{-1/2} .$$

Proof. Let $\theta \in \Theta$. Using Jensen's inequality with the convex function $t \mapsto (t+1)^{-1/2}$ on $\mathbb{R}_+^*$, we have

$$\int_\Theta \{\rho_\Theta^2(\theta,\nu)/2+1\}^{-1/2}\pi(\mathrm{d}\nu) \ge \{f_\pi(\theta)+1\}^{-1/2} . \tag{S61}$$

However, using the triangle and Hölder's inequalities, we have for any θ and $\nu \in \Theta$, $\rho_\Theta^2(\theta,\nu)/2 \le \rho_\Theta^2(\theta,\theta_\pi^\star) + \rho_\Theta^2(\theta_\pi^\star,\nu)$. Taking the integral with respect to π, by **MD**5 we get $f_\pi(\theta) \le \rho_\Theta^2(\theta,\theta_\pi^\star) + 2f_\pi(\theta_\pi^\star)$. Lastly, combining this result with (S61) and using that the function $t \mapsto (t+1)^{-1/2}$ is non-increasing on $\mathbb{R}_+^*$ completes the proof. □

Lemma S13. *Assume* **A** 2 *and* **MD** 5*. Let* $\theta_\pi^\star$ *be the Riemannian barycenter of the probability measure* π*, i.e.* $\theta_\pi^\star = \operatorname{argmin}_\Theta f_\pi$ *where* f_π *is defined by* (16)*. Then, for any* $\theta_0 \in \Theta$*,*

$$Q_\eta V_1(\theta_0) \le V_1(\theta_0) - [\eta/(4C_\pi^{1/2})] D_\Theta^2(\theta_0, \theta_\pi^\star) + 2\eta^2(1+\kappa)\{1 + f_\pi(\theta_\pi^\star)\}(f_\pi(\theta_\pi^\star) + 2) \ ,$$

where V_1 *is defined in* (9) *with* $\delta \leftarrow 1$*,* $\theta^\star \leftarrow \theta_\pi^\star$*,* $C_\pi = 1 + 2f_\pi(\theta_\pi^\star)$ *and* $D_\Theta^2 : \Theta^2 \to [0,1]$ *is defined by* (14)*.*

Proof. Let $\theta_0 \in \Theta$, and consider

$$H_{\theta_0}(X) = (1/2)\mathrm{Exp}_{\theta_0}^{-1}\left(X^{(1)}\right) \Big/ \left\{\rho_\Theta^2\left(\theta_0, X^{(2)}\right)\Big/2 + 1\right\}^{1/2} \ ,$$

where $X^{(1)}, X^{(2)}$ are independent random variables with distribution π.

Let $\gamma : [0,1] \to \Theta$ be the geodesic curve defined by $\gamma : t \mapsto \mathrm{Exp}_{\theta_0}[t\eta H_{\theta_0}(X)]$. Using [2, Lemma 1] with γ and V_1, we get

$$\begin{aligned} V_1(\gamma(1)) &\le V_1(\theta_0) + \langle \operatorname{grad} V_1(\theta_0), \dot\gamma(0)\rangle_{\theta_0} + (L/2)\,\|\dot\gamma(0)\|_{\theta_0}^2 \\ &= V_1(\theta_0) + \eta\,\langle \operatorname{grad} V_1(\theta_0), H_{\theta_0}(X)\rangle_{\theta_0} + ((1+\kappa)\eta^2/2)\,\|H_{\theta_0}(X)\|_{\theta_0}^2 \ , \end{aligned} \tag{S62}$$

by Proposition 5. We now compute the expectation of the terms in (S62). Using that $(X^{(1)}, X^{(2)})$ are independent, we obtain

$$\mathbb{E}\left[\langle \operatorname{grad} V_1(\theta_0), H_{\theta_0}(X)\rangle_{\theta_0}\right] = (1/2)\left\langle \operatorname{grad} V_1(\theta_0), \mathbb{E}\left[\mathrm{Exp}_{\theta_0}^{-1}\left(X^{(1)}\right)\right]\mathbb{E}\left[\left\{\rho_\Theta^2\left(\theta_0, X^{(2)}\right)\Big/2 + 1\right\}^{-1/2}\right]\right\rangle_{\theta_0} \ .$$

Moreover, using (S10) and Lemmas S11 and S12 yields

$$\begin{aligned} &\mathbb{E}\left[\langle \operatorname{grad} V_1(\theta_0), H_{\theta_0}(X)\rangle_{\theta_0}\right] \\ &\quad = -(1/2)\left\{\rho_\Theta^2(\theta_0, \theta_\pi^\star) + 1\right\}^{-1/2}\mathbb{E}\left[\left\langle \mathrm{Exp}_{\theta_0}^{-1}(\theta_\pi^\star), \mathrm{Exp}_{\theta_0}^{-1}(X^{(1)})\right\rangle_{\theta_0}\right]\mathbb{E}\left[\left\{\rho_\Theta^2(\theta_0, X^{(2)})\Big/2 + 1\right\}^{-1/2}\right] \\ &\quad \le -(1/4)\rho_\Theta^2(\theta_0, \theta_\pi^\star)\left[\left\{\rho_\Theta^2(\theta_0, \theta_\pi^\star) + 1\right\}\left\{\rho_\Theta^2(\theta_0, \theta_\pi^\star) + 2f_\pi(\theta_\pi^\star) + 1\right\}\right]^{-1/2} \\ &\quad \le -(16C_\pi)^{-1/2} D_\Theta^2(\theta_0, \theta_\pi^\star) \ , \end{aligned} \tag{S63}$$

where $C_\pi = 1 + 2f_\pi(\theta_\pi^\star)$ and $D_\Theta^2 : \Theta^2 \to [0,1]$ is defined by (14). Looking to bound the expectation of the last term in (S62), we use that $\|\mathrm{Exp}_{\theta_0}^{-1}(X^{(1)})\|_{\theta_0} = \rho_\Theta(\theta_0, X^{(1)})$ and that $X^{(1)}$ has distribution π to obtain,

$$\begin{aligned} \mathbb{E}\left[\|H_{\theta_0}(X)\|_{\theta_0}^2\right] &= (1/4)\mathbb{E}\left[\rho_\Theta^2(\theta_0, X^{(1)})\right]\mathbb{E}\left[\left\{\rho_\Theta^2(\theta_0, X^{(2)})/2 + 1\right\}^{-1}\right] \\ &= (f_\pi(\theta_0)/2)\mathbb{E}\left[\left\{\rho_\Theta^2(\theta_0, X^{(2)})/2 + 1\right\}^{-1}\right] \ . \end{aligned} \tag{S64}$$

Denote by $M = \rho_\Theta(\theta_\pi^\star, \theta_0)/2$. We bound the expectation in (S64) using the event $\{\rho_\Theta(\theta_\pi^\star, X^{(2)}) \ge M\}$ and its complement. On $\{\rho_\Theta(\theta_\pi^\star, X^{(2)}) \ge M\}$, we use Markov's inequality with the increasing map $t \mapsto t^2/2 + 1$,

$$\begin{aligned} \mathbb{E}\left[\mathbb{1}_{[M,+\infty)}(\rho_\Theta(\theta_\pi^\star, X^{(2)}))/[\rho_\Theta^2(\theta_0, X^{(2)})/2 + 1]\right] &\le \mathbb{P}\left(\rho_\Theta(\theta_\pi^\star, X^{(2)}) \ge M\right) \\ &\le \left(\mathbb{E}\left[\rho_\Theta^2(\theta_\pi^\star, X^{(2)})\right]\Big/2 + 1\right)\Big/\left(M^2/2 + 1\right) \ . \end{aligned} \tag{S65}$$

On $\{\rho_\Theta(\theta_\pi^\star, X^{(2)}) < M\}$, using the triangle inequality, we have

$$\rho_\Theta(\theta_0, X^{(2)}) \ge |\rho_\Theta(\theta_0, \theta_\pi^\star) - \rho_\Theta(\theta_\pi^\star, X^{(2)})| \ge \rho_\Theta(\theta_0, \theta_\pi^\star) - M = M \ .$$

Then, we obtain

$$\mathbb{E}\left[\mathbb{1}_{[0,M)}(\rho_\Theta(\theta, X^{(2)}))/\{\rho_\Theta^2(\theta_0, X^{(2)})/2 + 1\}\right] \le 1/[M^2/2 + 1] \ . \tag{S66}$$

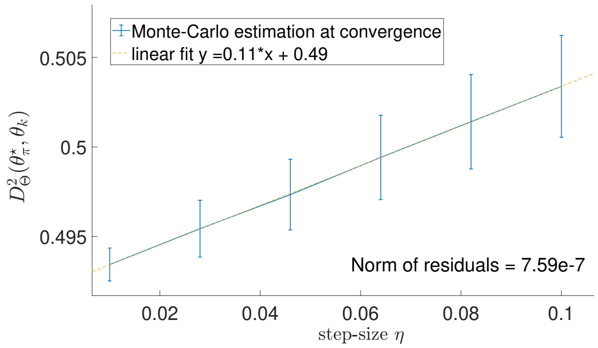

Figure S1: Monte Carlo approximations of the mean distance at convergence in Theorem 13

Adding (S65) and (S66) together and using the definition of M we obtain,

$$\mathbb{E}\left[\left\{\rho^2_{\Theta}\left(\theta_0, X^{(2)}\right)\Big/2+1\right\}^{-1}\right] \leq (f_\pi(\theta^\star_\pi)+2)/\left[\rho^2_{\Theta}(\theta^\star_\pi,\theta_0)/8+1\right] \ . \tag{S67}$$

Plugging (S67) in (S64), we get

$$\mathbb{E}\left[\|H_{\theta_0}(X)\|^2_{\theta_0}\right] \leq (f_\pi(\theta_0)/2)\,(f_\pi(\theta^\star_\pi)+2)\,/\,\left[\rho^2_{\Theta}(\theta^\star_\pi,\theta_0)/8+1\right] \ . \tag{S68}$$

Using the triangle and Hölder's inequalities, we have for any θ and $\nu \in \Theta$, $\rho^2_{\Theta}(\theta,\nu)/2 \leq \rho^2_{\Theta}(\theta,\theta^\star_\pi)+\rho^2_{\Theta}(\theta^\star_\pi,\nu)$. Taking the integral with respect to π, by **MD**5 we get $f_\pi(\theta) \leq \rho^2_{\Theta}(\theta,\theta^\star_\pi)+2f_\pi(\theta^\star_\pi)$. Combining this result and (S68), we obtain

$$\mathbb{E}\left[\|H_{\theta_0}(X)\|^2_{\theta_0}\right] \leq \{\rho^2_{\Theta}(\theta^\star_\pi,\theta_0)/2+f_\pi(\theta^\star_\pi)\}\,(f_\pi(\theta^\star_\pi)+2)\,/\,\left[\rho^2_{\Theta}(\theta^\star_\pi,\theta_0)/8+1\right] \leq 4\{1+f_\pi(\theta^\star_\pi)\}\,(f_\pi(\theta^\star_\pi)+2) \ .$$

Combining this result and (S63) in (S62) concludes the proof. □

Proof of Theorem 13. Let $\theta_0 \in \Theta, \eta > 0$ and $n \in \mathbb{N}$. Then, for any $k \in \{1,\ldots,n\}$, using Markov's property and Lemma S13 we have,

$$\begin{aligned}[\eta/(4C_\pi^{1/2})]\mathbb{E}\left[D^2_{\Theta}(\theta_{k-1},\theta^\star_\pi)\right] &= [\eta/(4C_\pi^{1/2})]\int_{\Theta} D^2_{\Theta}(\theta,\theta^\star_\pi)Q^{k-1}_{\eta}(\theta_0,\mathrm{d}\theta) \\ &\leq Q^{k-1}_{\eta}V_1(\theta_0) - Q^{k}_{\eta}V_1(\theta_0) + 2\eta^2(1+\kappa)(1+f_\pi(\theta^\star_\pi))(f_\pi(\theta^\star_\pi)+2) \ .\end{aligned}$$

Summing these inequalities for $k \in \{1,\ldots,n\}$ implies that

$$[\eta/(4C_\pi^{1/2})]\sum_{k=0}^{n-1}\mathbb{E}\left[D^2_{\Theta}(\theta_k,\theta^\star_\pi)\right] \leq V_1(\theta_0) - Q^n_{\eta}V_1(\theta_0) + 2n\eta^2(1+\kappa)(1+f_\pi(\theta^\star_\pi))(f_\pi(\theta^\star_\pi)+2) \ .$$

Finally, dividing both sides by $[n\eta/(4C_\pi^{1/2})]$ and using that V_1 is a non-negative function, we obtain

$$n^{-1}\sum_{k=0}^{n-1}\mathbb{E}\left[D^2_{\Theta}(\theta_k,\theta^\star_\pi)\right] \leq 2V_1(\theta_0)C_\pi^{1/2}\Big/(\eta n) + 2\eta(1+\kappa)(f_\pi(\theta^\star_\pi)+1)(f_\pi(\theta^\star_\pi)+2)(2f_\pi(\theta^\star_\pi)+1)^{-1/2} \ .$$

Which concludes the proof by setting $B_\pi = (1+\kappa)(f_\pi(\theta^\star_\pi)+1)(f_\pi(\theta^\star_\pi)+2)(2f_\pi(\theta^\star_\pi)+1)^{-1/2}$. □

Similarly to Figure 2, Figure S1 illustrates Theorem 7. To this end, 1000 replications of the experiment derived for Figure 3 are performed, obtaining $\{(\theta_n^{(i)}) : i \in \{1,\ldots,1000\}\}$ for $n = \lceil 50/\eta \rceil$ and $\eta \in \{1, 2.8, 4.6, 6.4, 8.2, 10\} \times 10^{-2}$. We estimate, with these samples, the mean and the variance of $D^2_{\Theta}(\theta,\theta^\star_\pi)$, for θ following the stationary distribution μ^η. We observe that the mean and variance are both linear w.r.t. the step-size η, indicating that the iterates of the SA scheme remain in a neighborhood of diameter $\mathcal{O}(\eta^{1/2})$ to the ground truth.

Even though the setting of this experiment goes beyond the assumptions of Theorem 7, it suggests that such a result may be applicable also in the setting of Theorem 13. The proof of such a result is left for future work.

S6 Background on Markov chain theory and Riemannian geometry

We give here some useful definitions and results that are used throughout the paper.

S6.1 Markov chain notions

We refer to [3] for a general introduction to Markov chains in general state space. Let $(\mathsf{Y}, \mathcal{Y})$ be a measurable state space and P be a Markov kernel on $\mathsf{Y} \times \mathcal{Y}$. Consider for any $y \in \mathsf{Y}$, the distribution $\mathbb{P}_y$ of the canonical Markov chain $(Y_n)_{n\in\mathbb{N}}$ corresponding to P and starting from y on the canonical space $(\mathsf{Y}^{\mathbb{N}}, \mathcal{Y}^{\otimes\mathbb{N}})$. Denote by $\mathbb{E}_y$ the corresponding expectation.

Denote for any $\mathsf{A} \in \mathcal{Y}$, $\tau_{\mathsf{A}} = \inf\{l \geq 1 : Y_l \in \mathsf{A}\}$ and $N_{\mathsf{A}} = \sum_{l=1}^{+\infty} \mathbb{1}_{\{\mathsf{A}\}}(Y_l)$.

We say that $(Y_n)_{n\in\mathbb{N}}$ is ψ-irreducible if there exists a measure ψ on $\mathcal{Y}$ such that whenever $\psi(\mathsf{A}) > 0$, we have $\mathbb{P}_y(\tau_{\mathsf{A}} < \infty) > 0$ for any $y \in \mathsf{Y}$. Moreover, a set $\mathsf{A} \in \mathcal{Y}$ is called Harris-recurrent if $\mathbb{P}_y(N_{\mathsf{A}} = \infty) = 1$ for any $y \in \mathsf{A}$. Finally, a chain $(Y_n)_{n\in\mathbb{N}}$ is called Harris-recurrent if it is ψ-irreducible and every set $\mathsf{A} \in \mathcal{Y}$ such that $\psi(\mathsf{A}) > 0$ is Harris-recurrent.

Let $\overline{V} : \mathsf{Y} \to [1, +\infty)$. We say that P is $\overline{V}$-uniformly geometrically ergodic if there exist $\rho \in [0, 1)$ and $C \geq 0$ such that for any $y \in \mathsf{Y}$ and $k \in \mathbb{N}$, $\left\|\delta_y P^k - \mu\right\|_{\overline{V}} \leq C\rho^k \overline{V}(y)$, where $\|\cdot\|_{\overline{V}}$ is defined for two probability measures ν_1, ν_2 on $(\mathsf{Y}, \mathcal{Y})$ by $\|\nu_1 - \nu_2\|_{\overline{V}} = \sup\{|\nu_1(g) - \nu_2(g)| : \sup_{\mathsf{Y}}\{|g| / \overline{V}\} \leq 1\}$.

S6.2 Useful results from Riemannian geometry

We now give definitions and auxiliary results related to tensor fields along curves, their derivatives, and Taylor expansions on Riemannian manifolds.

Let M be a smooth manifold with or without boundary. Given a smooth curve $\gamma : I \to \mathsf{M}$ defined on an interval I, and any $k, l \in \mathbb{N}$, a (k, l)-tensor field along γ is a continuous map $F : I \to \mathrm{T}^{(k,l)}\mathsf{TM}$, such that $F(t) \in \mathrm{T}^{(k,l)}(\mathrm{T}_{\gamma(t)}\mathsf{M})$ for any $t \in I$, where $\mathrm{T}^{(k,l)}\mathsf{TM}$ is the bundle of (k, l)-tensors on M, see e.g. [1, Appendix B]. A vector field Y along γ is a $(1, 0)$-tensor field, in which case for any $t \in I$, $Y(t)$ is just a tangent vector in $\mathrm{T}_{\gamma(t)}\mathsf{M}$. We say that a tensor field F along γ is extendible if there exists a tensor field $\tilde{F}$ defined on a neighborhood of $\gamma(I)$ such that $F = \tilde{F} \circ \gamma$.

We let $\mathfrak{X}^{k,l}(\gamma)$ denote the set of smooth (k, l)-tensor fields along γ, and $\mathfrak{X}(\gamma) = \mathfrak{X}^{1,0}(\gamma)$ denote the set of smooth vector fields along γ. In particular, $\mathfrak{X}^{0,0}(\gamma)$ is the set of smooth functions $g : I \to \gamma(I) \times \mathbb{R}$ such that for any $t \in I$, $g(t) = (\gamma(t), f(t))$ for some smooth function $f : I \to \mathbb{R}$ and therefore can be identified with the set of smooth functions $f : I \to \mathbb{R}$. In the sequel, we adopt if no confusion is possible this identification. We extend to tensor fields along γ the following definition of the trace on tensors. For any (k, l)-tensor T, we denote by $\mathrm{Tr}_{\square,\triangle}(T)$ the $(k-1, l-1)$-tensor with component of index $(i_1, \ldots, i_{k-1}, j_1, \ldots, j_{l-1})$, given by $\sum_{m=1}^{d} T^{j_1,\ldots,j_{\triangle-1},m,j_{\triangle},\ldots,j_{l-1}}_{i_1,\ldots,i_{\square-1},m,i_{\square},\ldots,i_{k-1}}$. In particular, for any $\omega \in \mathfrak{X}^{0,1}(\gamma), Y \in \mathfrak{X}(\gamma)$,

$$\mathrm{Tr}_{(1,1)}(\omega \otimes Y) = \omega(Y) .$$

Also, for any $F \in \mathfrak{X}^{k,l}(\gamma)$, any $\omega^1, \ldots, \omega^{k_0} \in \mathfrak{X}^{0,1}(\gamma)$ and $Y_1, \ldots, Y_{l_0} \in \mathfrak{X}(\gamma)$, with $k_0 \leq k, l_0 \leq l$, denote by $[F : \omega^1 \otimes \cdots \otimes \omega^{k_0} \otimes Y_1 \otimes \cdots \otimes Y_{l_0}]$, the $(k - k_0, l - l_0)$ smooth tensor field along γ defined by the induction:

$$[F : \omega^{\otimes 1:i}] = \mathrm{Tr}_{(1,l+1)}([F : \omega^{\otimes 1:(i-1)}] \otimes \omega^i) \tag{S69}$$

$$[F : \omega^{\otimes 1:k_0} \otimes Y_{\otimes 1:j}] = \mathrm{Tr}_{(k-k_0+1,1)}([F : \omega^{\otimes 1:k_0} \otimes Y_{\otimes 1:(j-1)}] \otimes Y_j) , \tag{S70}$$

setting $\omega^{\otimes 1:i} = \omega^1 \otimes \cdots \otimes \omega^i$, $Y_{\otimes 1:j} = Y_1 \otimes \cdots \otimes Y_j$. Note that for any $\omega^{k-k_0+1}, \ldots, \omega^k \in \mathfrak{X}^{0,1}(\gamma)$ and $Y_{l-l_0+1}, \ldots, Y_l \in \mathfrak{X}(\gamma)$,

$$[F : \omega^{\otimes 1:k_0} \otimes Y_{\otimes 1:l_0}](\omega^{k-k_0+1}, \ldots, \omega^k, Y_{l-l_0+1}, \ldots, Y_l) = F(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_l) . \tag{S71}$$

Proposition S14. *Let M be a smooth manifold with or without border, ∇ be a connection on $\mathrm{T}\mathsf{M}$ and $\gamma : I \to \mathsf{M}$ a smooth curve defined on an interval I. Then, for any $k, l \in \mathbb{N}$, ∇ determines an operator $\mathrm{D}_t : \mathfrak{X}^{k,l}(\gamma) \to \mathfrak{X}^{k,l}(\gamma)$, satisfying the following conditions.*

(a) On $\mathfrak{X}(\gamma)$, D_t is the usual covariant derivative along γ, see [1, Theorem 4.24].
(b) On $\mathfrak{X}^{0,0}(\gamma)$, D_t is the usual derivative for real functions, i.e. for any $f \in \mathfrak{X}^{0,0}(\gamma)$, $\mathrm{D}_t f = \mathrm{d}f/\mathrm{d}t$.
(c) For any $F \in \mathfrak{X}^{k,l}(\gamma)$, any $\omega^1, \ldots, \omega^k \in \mathfrak{X}^{0,1}(\gamma)$ and any $Y_1, \ldots, Y_l \in \mathfrak{X}(\gamma)$,

$$\begin{aligned}(\mathrm{D}_t F)\left(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_l\right) = \frac{\mathrm{d}}{\mathrm{d}t}&\left[F\left(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_l\right)\right] \\ &- \sum_{i=1}^{k} F\left(\omega^1, \ldots, \omega^{i-1}, \mathrm{D}_t\omega^i, \omega^{i+1}, \ldots, \omega^k, Y_1, \ldots, Y_l\right) \\ &- \sum_{j=1}^{l} F\left(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_{j-1}, \mathrm{D}_t Y_j, Y_{j+1}, \ldots, Y_l\right) \,. \end{aligned} \tag{S72}$$

In particular, D_t satisfies these additional properties.

(i) D_t satisfies the product rule, i.e. for any $f \in \mathfrak{X}^{0,0}(\gamma), F \in \mathfrak{X}^{k,l}(\gamma)$,

$$\mathrm{D}_t(fF) = \left(\frac{\mathrm{d}}{\mathrm{d}t} f\right) F + f \mathrm{D}_t F \,.$$

(ii) For any $k_1, l_1, k_2, l_2 \in \mathbb{N}$, and any $F \in \mathfrak{X}^{k_1,l_1}(\gamma), G \in \mathfrak{X}^{k_2,l_2}(\gamma)$,

$$\mathrm{D}_t(F \otimes G) = \mathrm{D}_t F \otimes G + F \otimes \mathrm{D}_t G \,.$$

(iii) For any positive integers $k_0 \leq k, l_0 \leq l$, $F \in \mathfrak{X}^{k,l}(\gamma)$,

$$\mathrm{D}_t\left\{\mathrm{Tr}_{(k_0,l_0)}(F)\right\} = \mathrm{Tr}_{(k_0,l_0)}\left(\mathrm{D}_t F\right) \,.$$

(iv) Let $F \in \mathfrak{X}^{k,l}$ be an extendible tensor field, i.e., such that there exists a (k,l)-tensor field $\tilde{F}$ defined on a neighborhood of $\gamma(I)$ satisfying for any $t \in I$, $F(t) = \tilde{F}(\gamma(t))$. Then, for any $t \in I$,

$$\mathrm{D}_t F(t) = \nabla_{\dot{\gamma}(t)} \tilde{F}(\gamma(t)) \,.$$

Finally, if $\tilde{\mathrm{D}}_t : \mathfrak{X}^{k,l}(\gamma) \to \mathfrak{X}^{k,l}(\gamma)$ is another operator satisfying (a),(b),(i),(ii) and (iii), then $\mathrm{D}_t = \tilde{\mathrm{D}}_t$.

Proof. Let $k, l \in \mathbb{N}$. Note first that (a)-(b) and (S72) define $\mathrm{D}_t F$ for any $F \in \mathfrak{X}^{k,l}(\gamma)$, setting for any $\omega \in \mathfrak{X}^{0,1}(\gamma)$ and $Y \in \mathfrak{X}(\gamma)$,

$$[\mathrm{D}_t\omega](Y) = \mathrm{d}\left[\omega(Y)\right]/\mathrm{d}t - \omega(\mathrm{D}_t Y) \,. \tag{S73}$$

We now show that $\mathrm{D}_t F \in \mathfrak{X}^{k,l}$, which will imply that $\mathrm{D}_t : \mathfrak{X}^{k,l} \to \mathfrak{X}^{k,l}$. Second, we establish that (i)-(ii)-(iii)-(iv) are satisfied. We conclude the proof by proving uniqueness of D_t.

Using [1, Lemma B.6], to show that $\mathrm{D}_t F \in \mathfrak{X}^{k,l}$ it is enough to prove that $\mathrm{D}_t F$ is multilinear over $\mathfrak{X}^{0,0}(\gamma)$. For that, we start proving (i) on $\mathfrak{X}^{0,1}(\gamma)$. Let $\omega \in \mathfrak{X}^{0,1}(\gamma), f \in \mathfrak{X}^{0,0}(\gamma)$ and $Y \in \mathfrak{X}(\gamma)$, then by (S73),

$$[\mathrm{D}_t(f\omega)](Y) = \mathrm{d}\left[f\omega(Y)\right]/\mathrm{d}t - f\omega\left(\mathrm{D}_t Y\right) = \left[\mathrm{d}f/\mathrm{d}t\right]\omega(Y) + f\left[\mathrm{D}_t\omega\right](Y) \,, \tag{S74}$$

which proves (i) on $\mathfrak{X}^{0,1}(\gamma)$. Now, let $k, l \in \mathbb{N}, F \in \mathfrak{X}^{k,l}(\gamma), \omega^1, \ldots, \omega^k \in \mathfrak{X}^{0,1}(\gamma), Y_1, \ldots, Y_l \in \mathfrak{X}(\gamma)$. Let $f \in \mathfrak{X}^{0,0}(\gamma)$ and $k_0 \in \mathbb{N}^*, k_0 \leq k$. We have, using the multilinearity of F over $\mathfrak{X}^{0,0}(\gamma)$, the definition of D_t (S72), and (S74)

$$\begin{aligned}&[\mathrm{D}_t F]\left(\omega^1, \ldots, \omega^{k_0-1}, f\omega^{k_0}, \omega^{k_0+1}, \ldots, \omega^k, Y_1, \ldots, Y_l\right) \\ &\quad= \frac{\mathrm{d}}{\mathrm{d}t}\left[F\left(\omega^1, \ldots, \omega^{k_0-1}, f\omega^{k_0}, \omega^{k_0+1}, \ldots, \omega^k, Y_1, \ldots, Y_l\right)\right] \\ &\qquad\qquad - \sum_{i=1, i\neq k_0}^{k} fF\left(\omega^1, \ldots, \omega^{i-1}, \mathrm{D}_t\omega^i, \omega^{i+1}, \ldots \omega^k, Y_1, \ldots, Y_l\right) \\ &\qquad\qquad - F\left(\omega^1, \ldots, \omega^{k_0-1}, \mathrm{D}_t(f\omega^{k_0}), \omega^{k_0+1}, \ldots, \omega^k, Y_1, \ldots, Y_l\right)\end{aligned}$$

$$
\begin{aligned}
&- \sum_{j=1}^{l} f F\left(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_{j-1}, \mathrm{D}_t Y_j, Y_{j+1}, \ldots, Y_l\right) \\
&= \left[\frac{\mathrm{d}}{\mathrm{d}t} f\right] \left\{F\left(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_k\right) - F\left(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_k\right)\right\} \\
&\qquad + f\left[\mathrm{D}_t F\right]\left(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_l\right) \\
&= f\left[\mathrm{D}_t F\right]\left(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_l\right) \ .
\end{aligned}
$$

The same arguments apply if we replace Y_{l_0} with fY_{l_0}, for some $l_0 \leq l$. Thus, using [1, Lemma B.6], $\mathrm{D}_t F \in \mathfrak{X}^{k,l}$.

Next, regarding (i), using the definition of D_t,

$$
\begin{aligned}
\left[\mathrm{D}_t f F\right]\left(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_l\right) = &\left[\frac{\mathrm{d}}{\mathrm{d}t} f\right] F\left(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_l\right) \\
&+ f\left[\mathrm{D}_t F\right]\left(\omega^1, \ldots, \omega^k, Y_1, \ldots, Y_l\right) \ ,
\end{aligned}
$$

thus proving (i). Moreover, we prove (ii). Let $k_1, l_1, k_2, l_2 \in \mathbb{N}$ and $F \in \mathfrak{X}^{k_1,l_1}(\gamma), G \in \mathfrak{X}^{k_2,l_2}(\gamma)$, $\omega^1, \ldots, \omega^{k_1+k_2} \in \mathfrak{X}^{0,1}(\gamma), Y_1, \ldots, Y_{l_1+l_2} \in \mathfrak{X}(\gamma)$. Setting

$$
f = F(\omega^1, \ldots, \omega^{k_1}, Y_1, \ldots, Y_{l_1}) \text{ and } g = G(\omega^{k_1+1}, \ldots, \omega^{k_1+k_2}, Y_{l_1+1}, \ldots, Y_{l_1+l_2}) \ ,
$$

we have

$$
\begin{aligned}
&\left[\mathrm{D}_t (F \otimes G)\right]\left(\omega^1, \ldots, \omega^{k_1+k_2}, Y_1, \ldots, Y_{l_1+l_2}\right) \\
&\quad = \frac{\mathrm{d}}{\mathrm{d}t}\left[fg\right] - \left[\sum_{i=1}^{k_1} F\left(\omega^1, \ldots, \omega^{i-1}, \mathrm{D}_t \omega^i, \omega^{i+1}, \ldots, \omega^{k_1}, Y_1, \ldots, Y_{l_1}\right)\right. \\
&\quad\quad \left. + \sum_{j=1}^{l_1} F\left(\omega^1, \ldots, \omega^{k_1}, Y_1, \ldots, Y_{j-1}, \mathrm{D}_t Y_j, Y_{j+1}, \ldots, Y_{l_1}\right)\right] g \\
&\quad\quad - f\left[\sum_{i=1}^{k_2} G\left(\omega^{k_1+1}, \ldots, \omega^{k_1+i-1}, \mathrm{D}_t \omega^{k_1+i}, \omega^{k_1+i+1}, \ldots, \omega^{k_1+k_2}, Y_{l_1+1}, \ldots, Y_{l_1+l_2}\right)\right. \\
&\quad\quad \left. + \sum_{j=1}^{l_2} G\left(\omega^{k_1+1}, \ldots, \omega^{k_1+k_2}, Y_{l_1+1}, \ldots, Y_{l_1+j-1}, \mathrm{D}_t Y_{l_1+j}, Y_{l_1+j+1}, \ldots, Y_{l_1+l_2}\right)\right] \\
&\quad = \left[\mathrm{D}_t F\right]\left(\omega^1, \ldots, \omega^{k_1}, Y_1, \ldots, Y_{l_1}\right) g + f\left[\mathrm{D}_t G\right]\left(\omega^{k_1+1}, \ldots, \omega^{k_1+k_2}, Y_{l_1+1}, \ldots, Y_{l_1+l_2}\right) \\
&\quad = \left[\mathrm{D}_t F \otimes G + F \otimes \mathrm{D}_t G\right]\left(\omega^1, \ldots, \omega^{k_1+k_2}, Y_1, \ldots, Y_{l_1+l_2}\right) \ ,
\end{aligned}
$$

which proves (ii). Furthermore, to prove (iii), let $t_0 \in I$ and $(\mathbf{b}_i)_{i \in \{1,\ldots,d\}}$ be a basis of $\mathrm{T}_{\gamma(t_0)}\Theta$. Using (a) and [1, Theorem 4.32], define for any $i \in \{1, \ldots, d\}$ and $t \in I$,

$$
e_i(t) = \mathrm{T}^{\gamma}_{t_0,t} \mathbf{b}_i \ ,
$$

where $\mathrm{T}^{\gamma}_{t_0,t}$ denotes the parallel transport map along γ from $\mathrm{T}_{\gamma(t_0)}\Theta$ to $\mathrm{T}_{\gamma(t)}\Theta$. As the parallel transport map is an isomorphism, $(e_i(t))_{i \in \{1,\ldots,d\}}$ is a basis of $\mathrm{T}_{\gamma(t)}\Theta$, for any $t \in I$. Therefore the family of smooth vector fields $(e_i)_{i \in \{1,\ldots,d\}}$ is a parallel frame along γ (with respect to ∇). Denote $(\varepsilon^j)_{j \in \{1,\ldots,d\}}$ its dual coframe. Using (S73) on $Y = e_i, \omega = \varepsilon^j$, for any $i, j \in \{1, \ldots, d\}$, shows that the coframe $(\varepsilon^j)_{j \in \{1,\ldots,d\}}$ is parallel along γ. Note that for $(e_i)_{i \in \{1,\ldots,d\}}$ and $(\varepsilon^j)_{j \in \{1,\ldots,d\}}$ to be well defined, we have used ∇, as well as the operator D_t on $\mathfrak{X}(\gamma)$ and $\mathfrak{X}^{0,1}(\gamma)$.

Let $k, l \in \mathbb{N}^*$ such that $k_0 \leq k, l_0 \leq l$, and let $F \in \mathfrak{X}^{k,l}(\gamma)$. There exist a family of functions $\{F^{j_1,\ldots,j_l}_{i_1,\ldots,i_k} \in \mathfrak{X}^{0,0}(\gamma) : i_1, \ldots, i_k, j_1, \ldots, j_l \in \{1, \ldots, d\}\}$ such that

$$
F = \sum_{i_1,\ldots,i_k=1}^{d} \sum_{j_1,\ldots,j_l=1}^{d} F^{j_1,\ldots,j_l}_{i_1,\ldots,i_k} \bigotimes_{\triangle=1}^{k} e_{i_\triangle} \bigotimes_{\square=1}^{l} \varepsilon^{j_\square} \ .
$$

Since the frame and its dual coframe are parallel along γ, for any $i \in \{1,\dots,d\}$ $\mathrm{D}_t e_i = 0$ and $\mathrm{D}_t \varepsilon^i = 0$. Combining this fact with (i) and (ii) gives

$$\mathrm{D}_t F = \sum_{i_1,\dots,i_k=1}^{d} \sum_{j_1,\dots,j_l=1}^{d} \left[\frac{\mathrm{d}}{\mathrm{d}t} F^{j_1,\dots,j_l}_{i_1,\dots,i_k}\right] \bigotimes_{\triangle=1}^{k} e_{i_\triangle} \bigotimes_{\square=1}^{l} \varepsilon^{j_\square} \; . \tag{S75}$$

Let $k_0, l_0 \in \mathbb{N}^*$ such that $k_0 \leq k, l_0 \leq l$, then by definition of $\mathrm{Tr}_{(k_0,l_0)}$, for any $i_1,\dots,i_{k-1},j_1,\dots,j_{l-1} \in \{1,\dots,d\}$,

$$\mathrm{Tr}_{(k_0,l_0)}(F)^{j_1,\dots,j_{l-1}}_{i_1,\dots,i_{k-1}} = \sum_{m=1}^{d} F^{j_1,\dots,j_{l_0-1},m,j_{l_0},\dots,j_{l-1}}_{i_1,\dots,i_{k_0-1},m,i_{k_0},\dots,i_{k-1}} \; . \tag{S76}$$

We remind the reader that $\mathrm{Tr}_{(k_0,l_0)}(F)$ does not depend on the choice of coordinates [1, Appendix B]. Thus, using (S75) and (S76), we have

$$\begin{aligned}
\mathrm{D}_t \left[\mathrm{Tr}_{(k_0,l_0)}(F)\right] &= \sum_{i_1,\dots,i_{k-1}=1}^{d} \sum_{j_1,\dots,j_{l-1}=1}^{d} \frac{\mathrm{d}}{\mathrm{d}t} \left[\mathrm{Tr}_{(k_0,l_0)}(F)^{j_1,\dots,j_{l-1}}_{i_1,\dots,i_{k-1}}\right] \bigotimes_{\triangle=1}^{k-1} e_{i_\triangle} \bigotimes_{\square=1}^{l-1} \varepsilon^{j_\square} \\
&= \sum_{i_1,\dots,i_{k-1}=1}^{d} \sum_{j_1,\dots,j_{l-1}=1}^{d} \sum_{m=1}^{d} \frac{\mathrm{d}}{\mathrm{d}t} F^{j_1,\dots,j_{l_0-1},m,j_{l_0},\dots,j_{l-1}}_{i_1,\dots,i_{k_0-1},m,i_{k_0},\dots,i_{k-1}} \bigotimes_{\triangle=1}^{k-1} e_{i_\triangle} \bigotimes_{\square=1}^{l-1} \varepsilon^{j_\square} \\
&= \mathrm{Tr}_{(k_0,l_0)}\left(\mathrm{D}_t F\right) \; ,
\end{aligned}$$

thus proving (iii).

To prove (iv), first for any $f \in \mathfrak{X}^{(0,0)}(\gamma)$, extendible in $\tilde{f}$, we have by composition and definition of the covariant derivative, that for any $t \in [0,1]$,

$$(\mathrm{d}f/\mathrm{d}t)(t) = \mathrm{d}\tilde{f}_{\gamma(t)}(\dot{\gamma}(t)) = \nabla_{\dot{\gamma}(t)} \tilde{f}(\gamma(t)) \; . \tag{S77}$$

Also, using [1, Theorem 4.24-(iii)] gives (iv) for any $Y \in \mathfrak{X}(\gamma)$. Combining (S77), (S73), its counterpart for tensor fields defined over a manifold [1, Proposition 4.15-(a)] and (iv) over $\mathfrak{X}(\gamma)$, proves (iv) over $\mathfrak{X}^{(0,1)}(\gamma)$. Now, for any $k,l \in \mathbb{N}$, using (iv) over $\mathfrak{X}(\gamma)$ and $\mathfrak{X}^{(0,1)}(\gamma)$ combined with (S72) and its counterpart for tensor fields defined over a manifold [1, Equation (4.12)] gives (iv) over $\mathfrak{X}^{(k,l)}(\gamma)$.

Finally, we address uniqueness. Suppose now that $\tilde{\mathrm{D}}_t$ is an operator on $\mathfrak{X}^{k,l}(\gamma)$ that satisfies (a),(b),(i),(ii) and (iii). First, (a) and (b) show that D_t and $\tilde{\mathrm{D}}_t$ coincide on $\mathfrak{X}^{0,0}(\gamma)$ and $\mathfrak{X}(\gamma)$. Second, for any $Y \in \mathfrak{X}(\gamma), \omega \in \mathfrak{X}^{0,1}(\gamma)$, writing $\omega(Y) = \mathrm{Tr}_{(1,1)}(Y \otimes \omega)$ and using (iii) gives

$$\tilde{\mathrm{D}}_t \omega = \mathrm{d}[\omega(Y)]/\mathrm{d}t - \omega(\tilde{\mathrm{D}}_t Y) = \mathrm{D}_t \omega \; ,$$

using (S73). Thus, $\tilde{\mathrm{D}}_t$ and D_t also agree on $\mathfrak{X}^{0,1}(\gamma)$. Therefore, the frame $(e_i)_{i\in\{1,\dots,d\}}$ and its dual coframe $(\varepsilon^j)_{j\in\{1,\dots,d\}}$ are also parallel with respect to $\tilde{\mathrm{D}}_t$ along γ. Let $F \in \mathfrak{X}^{k,l}(\gamma)$, then using (i) and (ii) shows that (S75) holds for the operator $\tilde{\mathrm{D}}_t$, proving that $\mathrm{D}_t F = \tilde{\mathrm{D}}_t F$. This concludes the proof.

□

Lemma S15. *Let* M *be a smooth manifold and* ∇ *be a connection on* $\mathrm{T}\mathsf{M}$. *Let* $\gamma : [0,1] \to \mathsf{M}$ *be a smooth curve and denote* D_t *the covariant derivative operator along* γ *associated with* ∇, *defined in Proposition S14. Let* $F \in \mathfrak{X}^{k,l}(\gamma)$, $\omega^1,\dots,\omega^{k_0} \in \mathfrak{X}^{0,1}(\gamma)$ *and* $Y_1,\dots,Y_{l_0} \in \mathfrak{X}(\gamma)$, *with* $k_0 \leq k, l_0 \leq l$. *Then, we have*

$$\begin{aligned}
\mathrm{D}_t \left([F : \omega^{\otimes 1:k_0} \otimes Y_{\otimes 1:l_0}]\right) &= [\mathrm{D}_t F : \omega^{\otimes 1:k_0} \otimes Y_{\otimes 1:l_0}] \\
&\quad + \sum_{i=1}^{k_0} [F : \omega^{\otimes 1:(i-1)} \otimes \mathrm{D}_t \omega^i \otimes \omega^{(i+1):k_0} \otimes Y_{\otimes 1:l_0}] \\
&\quad + \sum_{j=1}^{l_0} [F : \omega^{\otimes 1:k_0} \otimes Y_{\otimes 1:(j-1)} \otimes \mathrm{D}_t Y_j \otimes Y_{\otimes (j+1):l_0}] \; .
\end{aligned} \tag{S78}$$

Proof. Let F be a smooth (k,l)-tensor field along γ. We show (S78) by induction. Following the recursive definition of the contraction in (S69), we prove it by induction on $k_0 \in \mathbb{N}^*, k_0 \leq k$, for any $\omega^1, \ldots, \omega^{k_0} \in \mathfrak{X}^{0,1}(\gamma)$.

The case $k_0 = 1$ follows from Proposition S14-(ii) and (iii), combined with the definition in (S69),

$$\begin{aligned} \mathrm{D}_t \left[F : \omega^1\right] &= \mathrm{D}_t \operatorname{Tr}_{(1,l+1)}(F \otimes \omega^1) \\ &= \operatorname{Tr}_{(1,l+1)}(\mathrm{D}_t[F \otimes \omega^1]) \\ &= \operatorname{Tr}_{(1,l+1)}(\mathrm{D}_t F \otimes \omega^1 + F \otimes \mathrm{D}_t \omega^1) \\ &= \left[\mathrm{D}_t F : \omega^1\right] + \left[F : \mathrm{D}_t \omega^1\right] \ , \end{aligned}$$

where we have used the linearity of Tr. Now assume there exists $k_0 \in \{1, \ldots, k-1\}$ such that (S78) holds for any smooth 1 forms $\omega^1, \ldots, \omega^{k_0}$ and $l_0 = 0$. Moreover, consider any smooth 1 forms $\omega^1, \ldots, \omega^{k_0+1}$. Then, using the same arguments as for the case $k_0 = 1$ and the induction hypothesis, we obtain

$$\begin{aligned} \mathrm{D}_t \left[F : \omega^{\otimes 1:(k_0+1)}\right] &= \mathrm{D}_t \operatorname{Tr}_{(1,l+1)} \left(\left[F : \omega^{\otimes 1:k_0}\right] \otimes \omega^{k_0+1}\right) \\ &= \operatorname{Tr}_{(1,l+1)} \left(\mathrm{D}_t \left[F : \omega^{\otimes 1:k_0}\right] \otimes \omega^{k_0+1}\right) + \operatorname{Tr}_{(1,l+1)} \left(\left[F : \omega^{\otimes 1:k_0}\right] \otimes \mathrm{D}_t \omega^{k_0+1}\right) \\ &= \operatorname{Tr}_{(1,l+1)} \left(\left[\mathrm{D}_t F : \omega^{\otimes 1:k_0}\right] \otimes \omega^{k_0+1}\right) + \left[F : \omega^{\otimes 1:k_0} \otimes \mathrm{D}_t \omega^{k_0+1}\right] \\ &\quad + \sum_{i=1}^{k_0} \operatorname{Tr}_{(1,l+1)} \left(\left[F : \omega^{\otimes 1:(i-1)} \otimes \mathrm{D}_t \omega^i \otimes \omega^{\otimes (i+1):k_0}\right] \otimes \omega^{k_0+1}\right) \\ &= \left[\mathrm{D}_t F : \omega^{\otimes 1:(k_0+1)}\right] + \sum_{i=1}^{k_0+1} \left[F : \omega^{\otimes 1:(i-1)} \otimes \mathrm{D}_t \omega^i \otimes \omega^{\otimes (i+1):(k_0+1)}\right] \ . \end{aligned}$$

Subsequently, using the recursive definition of the contraction in (S70), we prove (S78) by induction on $l_0 \in \mathbb{N}^*, l_0 \leq l$ for any $k_0 \leq k$ and any $\omega^1, \ldots, \omega^{k_0} \in \mathfrak{X}^{0,1}(\gamma)$. Let $Y_1 \in \mathfrak{X}(\gamma)$. Then, using once again Proposition S14-(ii) and (iii), (S70), and (S78) in the case $l_0 = 0$ justified above, the case $l_0 = 1$ is proven as follows,

$$\begin{aligned} \mathrm{D}_t \left[F : \omega^{\otimes 1:k_0} \otimes Y_1\right] &= \operatorname{Tr}_{(k-k_0+1,1)} \left(\mathrm{D}_t \left\{\left[F : \omega^{\otimes 1:k_0}\right] \otimes Y_1\right\}\right) \\ &= \operatorname{Tr}_{(k-k_0+1,1)} \left(\left[\mathrm{D}_t F : \omega^{\otimes 1:k_0}\right] \otimes Y_1\right) + \left[F : \omega^{\otimes 1:k_0} \otimes \mathrm{D}_t Y_1\right] \\ &\quad + \sum_{i=1}^{k_0} \operatorname{Tr}_{(k-k_0+1,1)} \left(\left[F : \omega^{\otimes 1:(i-1)} \otimes \mathrm{D}_t \omega^i \otimes \omega^{\otimes (i+1):k_0}\right] \otimes Y_1\right) \\ &= \left[\mathrm{D}_t F : \omega^{\otimes 1:k_0} \otimes Y_1\right] + \sum_{i=1}^{k_0} \left[F : \omega^{\otimes 1:(i-1)} \otimes \mathrm{D}_t \omega^i \otimes \omega^{\otimes (i+1):k_0} \otimes Y_1\right] \\ &\quad + \left[F : \omega^{\otimes 1:k_0} \otimes \mathrm{D}_t Y_1\right] \ . \end{aligned}$$

Furthermore, assume there exists $l_0 \in \{1, \ldots, l-1\}$ such that (S78) holds for any $k_0 \leq k$, any $\omega^1, \ldots, \omega^{k_0} \in \mathfrak{X}^{0,1}(\gamma)$ and any $Y_1, \ldots, Y_{l_0} \in \mathfrak{X}(\gamma)$. Let $Y_1, \ldots, Y_{l_0+1} \in \mathfrak{X}(\gamma)$. Then using the same arguments as for the case $l_0 = 1$ and the induction hypothesis, we obtain

$$\begin{aligned} &\mathrm{D}_t \left[F : \omega^{\otimes 1:k_0} \otimes Y_{\otimes 1:(l_0+1)}\right] \\ &\quad = \operatorname{Tr}_{(k-k_0+1,1)} \left(\mathrm{D}_t \left\{\left[F : \omega^{\otimes 1:k_0} \otimes Y_{\otimes 1:l_0}\right] \otimes Y_{l_0+1}\right\}\right) \\ &\quad = \operatorname{Tr}_{(k-k_0+1,1)} \left(\left[\mathrm{D}_t F : \omega^{\otimes 1:k_0} \otimes Y_{\otimes 1:l_0}\right] \otimes Y_{l_0+1}\right) \\ &\qquad + \sum_{i=1}^{k_0} \operatorname{Tr}_{(k-k_0+1,1)} \left(\left[F : \omega^{\otimes 1:(i-1)} \otimes \mathrm{D}_t \omega^i \otimes \omega^{\otimes (i+1):k_0} \otimes Y_{\otimes 1:l_0}\right] \otimes Y_{l_0+1}\right) \\ &\qquad + \sum_{j=1}^{l_0} \operatorname{Tr}_{(k-k_0+1,1)} \left(\left[F : \omega^{\otimes 1:k_0} \otimes Y_{\otimes 1:(j-1)} \otimes \mathrm{D}_t Y_j \otimes Y_{\otimes (j+1):l_0}\right] \otimes Y_{l_0+1}\right) \\ &\qquad + \operatorname{Tr}_{(k-k_0+1,1)} \left(\left[F : \omega^{\otimes 1:k_0} \otimes Y_{\otimes 1:l_0}\right] \otimes \mathrm{D}_t Y_{l_0+1}\right) \\ &\quad = \left[\mathrm{D}_t F : \omega^{\otimes 1:k_0} \otimes Y_{\otimes 1:(l_0+1)}\right] + \sum_{i=1}^{k_0} \left[F : \omega^{\otimes 1:(i-1)} \otimes \mathrm{D}_t \omega^i \otimes \omega^{\otimes (i+1):k_0} \otimes Y_{\otimes 1:(l_0+1)}\right] \end{aligned}$$

$$+\sum_{j=1}^{l_0+1}\left[F:\omega^{\otimes 1:k_0}\otimes Y_{\otimes 1:(j-1)}\otimes \mathrm{D}_t Y_j \otimes Y_{\otimes (j+1):(l_0+1)}\right] ,$$

which concludes the proof. □

Theorem S16. *Let* M *be a smooth manifold and* ∇ *be a connection on* TM*. Let* $\gamma : [0,1]\to\mathsf{M}$ *be a geodesic and* $Y:\mathsf{M}\to\mathsf{TM}$ *a smooth vector field. Then, for any* $t\in[0,1], n\in\mathbb{N}$*,*

$$\begin{aligned}\mathrm{T}^{\gamma}_{t0}Y(\gamma(t)) = \sum_{k_0=0}^{n}(t^{k_0}/k_0!)\;\nabla^{k_0}Y_{\gamma(0)}\left(\dot\gamma(0),\dots,\dot\gamma(0)\right)\\ +\int_0^t[(t-s)^n/n!]\mathrm{T}^{\gamma}_{s0}\nabla^{n+1}Y_{\gamma(s)}\left(\dot\gamma(s),\dots,\dot\gamma(s)\right)\mathrm{d}s ,\end{aligned} \tag{S79}$$

where $\mathrm{T}^{\gamma}_{t0}:\mathrm{T}_{\gamma(t)}\mathsf{M}\to\mathrm{T}_{\gamma(0)}\mathsf{M}$ *is the parallel transport map along* γ*, and the* $(1,k_0)$*-tensor field* $\nabla^{k_0}Y$ *is the total derivative of order* k_0 *of the* $(1,0)$*-tensor field* Y*.*

For a definition of the total covariant derivative, see [1, Proposition 4.15]. Also, in (S79), remark that even though $\dot\gamma$ is only a vector field along γ, and not a vector field, the value of a vector field $\nabla_X Y$ evaluated at $\theta\in\mathsf{M}$ only depends on $X(\theta)$ and on values of Y along smooth curves $\mathrm{c}:[0,1]\to\mathsf{M}$ satisfying $\mathrm{c}(0)=\theta$ and $\dot{\mathrm{c}}(0)=X(\theta)$; by [1, Proposition 4.26]. Therefore the expression $\nabla^{k_0}Y_{\gamma(t)}(\dot\gamma(t),\dots,\dot\gamma(t))$ in Theorem S16 is well defined for any $k_0\in\mathbb{N}, t\in[0,1]$.

Proof. Consider $\mathcal{V}:[0,1]\to\mathsf{M}$ the smooth vector field along γ and the function $\varphi:[0,1]\to\mathrm{T}_{\gamma(0)}\mathsf{M}$ defined by

$$\mathcal{V}=Y\circ\gamma \text{ and } \varphi : t\mapsto \mathrm{T}^{\gamma}_{t0}\mathcal{V}(t) .$$

Then we check by induction on $n\in\mathbb{N}^*$ that φ is n-times differentiable with derivative of order n given for any $t\in[0,1]$ by $\varphi^{(n)}(t)=\mathrm{T}^{\gamma}_{t0}[\mathrm{D}^n_t\mathcal{V}(t)]$ and $\mathrm{D}^n_t\mathcal{V}(t)=\nabla^n Y_{\gamma(t)}(\dot\gamma(t),\dots,\dot\gamma(t))$, where D_t is the covariant derivative operator along γ with respect to the connection ∇, defined in Proposition S14.

First, the case $n=1$ is a direct application of [1, Theorem 4.34, Theorem 4.24] since Y is an extension of $\mathcal{V}$. Assume now that the property holds for $n\in\mathbb{N}^*$. Then, for any $t_0,t\in[0,1], t\neq t_0$, we have

$$\left[\varphi^{(n)}(t)-\varphi^{(n)}(t_0)\right]\Big/(t-t_0) = \mathrm{T}^{\gamma}_{t_0 0}\left[\mathrm{T}^{\gamma}_{tt_0}\mathrm{D}^n_t\mathcal{V}(t)-\mathrm{D}^n_t\mathcal{V}(t_0)\right]\big/(t-t_0) .$$

Now [1, Theorem 4.34] ensures that the limit of the quantity above exists when $t\to t_0$ and in addition this limit is

$$\varphi^{(n+1)}(t_0)=\mathrm{T}^{\gamma}_{t_0 0}\mathrm{D}^{n+1}_t\mathcal{V}(t_0) ,$$

which shows that φ is $n+1$ times differentiable on $[0,1]$. We now show that for any $t\in[0,1]$, $\mathrm{D}^{n+1}_t\mathcal{V}(t)=\nabla^{n+1}Y_{\gamma(t)}(\dot\gamma(t),\dots,\dot\gamma(t))$. Using Lemma S15 on the smooth $(1,n)$-tensor field along γ $F=(\nabla^n Y)\circ\gamma$, taking $k_0=0$ and n times the vector field $\dot\gamma$, we have

$$\mathrm{D}_t\left[F:\dot\gamma\otimes\cdots\otimes\dot\gamma\right]=\left[\mathrm{D}_tF:\dot\gamma\otimes\cdots\otimes\dot\gamma\right] ,$$

since $\mathrm{D}_t\dot\gamma=0$ because γ is a geodesic. Also, by (S71), $[\mathrm{D}_tF:\dot\gamma\otimes\cdots\otimes\dot\gamma]=\mathrm{D}_tF(\dot\gamma,\dots,\dot\gamma)$. Finally, as $\nabla^n Y$ is an extension of F, using the induction hypothesis and the definition of the total derivative give for any $t\in[0,1]$,

$$\begin{aligned}\mathrm{D}^{n+1}_t\mathcal{V}(t)=\mathrm{D}_tF\left(\dot\gamma,\dots,\dot\gamma\right)(t)=\nabla_{\dot\gamma(t)}(\nabla^nY)_{\gamma(t)}\left(\dot\gamma(t),\dots,\dot\gamma(t)\right)\\ =\left(\nabla^{n+1}Y\right)_{\gamma(t)}\left(\dot\gamma(t),\dots,\dot\gamma(t)\right) ,\end{aligned}$$

concluding the induction.

Finally, (S79) is simply a consequence of Taylor's formula with integral remainder of the vectorial valued function φ identifying $\mathrm{T}_{\gamma(0)}\mathsf{M}$ with $\mathbb{R}^d$. □

Proposition S17. *Let* M *be a smooth manifold,* ∇ *be a symmetric connection defined over the smooth vector fields of* M*. For any smooth function* $f : \mathsf{M} \to \mathbb{R}$ *and any local coordinates* $(u_i)_{i\in\{1,\ldots,d\}}$*, we have*

$$\nabla \mathrm{Hess}\, f = \sum_{i,j,k=1}^{d} \left\{ \partial^3_{kij} f - \sum_{l=1}^{d} \left[\Gamma^l_{ij} \partial^2_{kl} f + \Gamma^l_{ki} \partial^2_{jl} f + \Gamma^l_{kj} \partial^2_{il} f \right] - \sum_{m=1}^{d} \partial_k \Gamma^m_{ij} \partial_m f + \sum_{l,m=1}^{d} \left[\Gamma^l_{kj} \Gamma^m_{il} + \Gamma^l_{ki} \Gamma^m_{lj} \right] \partial_m f \right\} \mathrm{d}u^i \otimes \mathrm{d}u^j \otimes \mathrm{d}u^k ,$$

where $(\Gamma^k_{ij})_{i,j,k\in\{1,\ldots,d\}}$ *are the Christoffel symbols in these local coordinates, the local frame and its dual coframe are denoted by* $(\partial u_i)_{i\in\{1,\ldots,d\}}$ *and* $(\mathrm{d}u^j)_{j\in\{1,\ldots,d\}}$*.*

Proof. Let $(u_i)_{i\in\{1,\ldots,d\}}$ be local coordinates. By [1, Example 4.22], in this chart, we have

$$\mathrm{Hess}\, f = \sum_{i,j=1}^{d} F_{ij} \mathrm{d}u^i \otimes \mathrm{d}u^j \text{ ,where for any } i,j \in \{1,\ldots,d\} ,\ F_{ij} = \partial^2_{ij} f - \sum_{m=1}^{d} \Gamma^m_{ij} \partial_m f . \tag{S80}$$

Applying [1, Proposition 4.18] on Hess f, we obtain that $\nabla \mathrm{Hess}\, f = \sum_{i,j,k=1}^{d} G_{ijk} \mathrm{d}u^i \otimes \mathrm{d}u^j \otimes \mathrm{d}u^k$, where for any $i,j,k \in \{1,\ldots,d\}$,

$$G_{ijk} = \partial_k F_{ij} - \sum_{l=1}^{d} \left(\Gamma^l_{kj} F_{il} + \Gamma^l_{ki} F_{lj} \right) .$$

Expanding the expression above using (S80) gives for any $i,j,k \in \{1,\ldots,d\}$,

$$G_{ijk} = \partial^3_{ijk} f - \sum_{m=1}^{d} \left(\partial_k \Gamma^m_{ij} \partial_m f + \Gamma^m_{ij} \partial^2_{km} f \right) - \sum_{l=1}^{d} \Gamma^l_{kj} \left(\partial^2_{il} f - \sum_{m=1}^{d} \Gamma^m_{il} \partial_m f \right) - \sum_{l=1}^{d} \Gamma^l_{ki} \left(\partial^2_{lj} f - \sum_{m=1}^{d} \Gamma^m_{lj} \partial_m f \right) .$$

The desired result is obtained by reordering this equation, which concludes the proof. □

References

[1] J. M. Lee. *Introduction to Riemannian Manifolds.* Springer International Publishing, 2019.

[2] A. Durmus, P. Jiménez, É. Moulines, S. Said, and H. T. Wai. Convergence analysis of Riemannian stochastic approximation schemes. *arXiv preprint arXiv:2005.13284*, 2020.

[3] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability.* Cambridge University Press, New York, NY, USA, 2nd edition, 2009.

[4] J. Jost. *Riemannian Geometry and Geometric Analysis.* Springer Universitat texts. Springer, 2005.

[5] K. T. Sturm. Probability Measures on Metric Spaces of Nonpositive Curvature. *Contemporary Mathematics*, 338, 01 2003.

[6] R. A. Horn and C. R. Johnson. *Topics in matrix analysis.* Cambridge university press, 1994.

[7] J. Kent. Time-reversible diffusions. *Adv. in Appl. Probab.*, 10(4):819–835, 1978.

[8] Nicolas Boumal. An introduction to optimization on smooth manifolds. Available online, Aug 2020.