Supplementary material: Intrinsic wavelet regression for curves of Hermitian positive definite matrices

Joris Chau* and Rainer von Sachs

Contents

7 Appendix I: Geometry of HPD matrices

The space of $(d \times d)$-dimensional Hermitian matrices together with matrix addition and scalar multiplication $(\mathbb{H}_{d\times d}, +, \cdot_S)$ is a real vector space and every finite-dimensional real vector space has a natural smooth manifold structure by considering a global coordinate chart induced by a basis of the real vector space. The space of $(d \times d)$-dimensional Hermitian positive definite (HPD) matrices is no longer a vector space due to the positive definite constraints, but it is an open subset of $\mathbb{H}_{d\times d}$ and as such it is also a smooth manifold, see e.g. do Carmo (1992).

Affine-invariant Riemannian metric For notational convenience, in the remainder of the supplemental document, we denote $\mathcal{M} := \mathbb{P}_{d\times d}$ for the space of $(d \times d)$-dimensional HPD matrices, an d^2-dimensional smooth manifold. For every $p \in \mathcal{M}$, the tangent space $T_p(\mathcal{M})$ can be identified by $\mathcal{H} := \mathbb{H}_{d\times d}$, the space of $(d \times d)$-dimensional Hermitian matrices. As detailed

*Corresponding author, joris.chau@openanalytics.be, Institute of Statistics, Biostatistics, and Actuarial Sciences (ISBA), Université catholique de Louvain, Voie du Roman Pays 20, B-1348, Louvain-la-Neuve, Belgium.

in Pennec et al. (2006), the Frobenius inner product on $\mathbb{H}_{d\times d}$ induces the affine-invariant Riemannian metric g_R on the manifold $\mathcal{M}$ given by the smooth family of inner products:

$$\langle h_1, h_2\rangle_p \quad = \quad \mathrm{Tr}((p^{-1/2} * h_1)(p^{-1/2} * h_2)), \qquad \forall\, p \in \mathcal{M}, \tag{7.1}$$

with notation as in the main document and $h_1, h_2 \in T_p(\mathcal{M})$. The Riemannian distance on $\mathcal{M}$ derived from the Riemannian metric is given by:

$$\delta_R(p_1, p_2) \quad = \quad \|\mathrm{Log}(p_1^{-1/2} * p_2)\|_F, \tag{7.2}$$

The mapping $x \mapsto a * x$ is an isometry for each invertible matrix $a \in \mathrm{GL}(d, \mathbb{C}) = \{A \in \mathbb{C}^{d\times d} \mid \det(A) \neq 0\}$, i.e., it is distance-preserving:

$$\delta_R(p_1, p_2) \quad = \quad \delta_R(a * p_1, a * p_2), \quad \forall\, a \in \mathrm{GL}(d, \mathbb{C}).$$

Geodesics By (Bhatia, 2009, Theorem 6.1.6 and Prop. 6.2.2), the Riemannian manifold $(\mathcal{M}, g_R)$, with g_R the affine-invariant metric, is geodesically complete, and the *geodesic* segment joining any two points $p_1, p_2 \in \mathcal{M}$ is unique and can be parametrized as,

$$\eta(p_1, p_2, t) \quad = \quad p_1^{1/2} * \left(p_1^{-1/2} * p_2\right)^t, \quad 0 \leq t \leq 1. \tag{7.3}$$

Exp- and Log-maps Since $(\mathcal{M}, g_R)$ is a geodesically complete manifold, the Hopf-Rinow Theorem says that for every $p \in \mathcal{M}$ the exponential map Exp_p and the logarithmic map Log_p are global diffeomorphisms with as domains $T_p(\mathcal{M})$ and $\mathcal{M}$ respectively. By (Pennec et al. (2006)), the exponential map $\mathrm{Exp}_p : T_p(\mathcal{M}) \to \mathcal{M}$ is given by,

$$\mathrm{Exp}_p(h) \quad = \quad p^{1/2} * \mathrm{Exp}\left(p^{-1/2} * h\right), \quad \forall\, h \in T_p(\mathcal{M}), \tag{7.4}$$

The logarithmic map $\mathrm{Log}_p : \mathcal{M} \to T_p(\mathcal{M})$ is given by the inverse exponential map:

$$\mathrm{Log}_p(q) \quad = \quad p^{1/2} * \mathrm{Log}\left(p^{-1/2} * q\right). \tag{7.5}$$

The Riemannian distance may now also be expressed in terms of the logarithmic map as:

$$\delta_R(p_1, p_2) \quad = \quad \|\mathrm{Log}_{p_1}(p_2)\|_{p_1} \quad = \quad \|\mathrm{Log}_{p_2}(p_1)\|_{p_2}, \quad \forall\, p_1, p_2 \in \mathcal{M}, \tag{7.6}$$

where $\|h\|_p := \langle h, h\rangle_p$ denotes the norm of $h \in T_p(\mathcal{M})$ induced by the affine-invariant Riemannian metric.

Parallel transport As outlined in Jeuris et al. (2012) among others, the covariant derivative at $p \in \mathcal{M}$ of a smooth vector field $Y \in \mathfrak{X}(\mathcal{M})$, with respect to a smooth vector field $X \in \mathfrak{X}(\mathcal{M})$ is given by:

$$(\nabla_{X_p} Y)_p \quad = \quad D(Y)(p)[X_p] - \frac{1}{2}(X_p p^{-1} Y_p + Y_p p^{-1} X_p). \tag{7.7}$$

Here, $X_p, Y_p \in T_p(\mathcal{M})$ denote the tangent vectors associated with the vector fields X, Y at $p \in \mathcal{M}$ and $D(Y)(p)[X_p] := \lim_{h\to 0}(Y(p + hX_p) - Y(p))/h$ is the classical Fréchet derivative of $Y(p)$, where $Y : \mathcal{M} \to T\mathcal{M}$ maps $p \in \mathcal{M}$ to the tangent vector $Y_p \in T_p(\mathcal{M})$. This connection

∇ is exactly the Levi-Civita connection on the Riemannian manifold $(\mathcal{M}, g_R)$, as it can be verified that it satisfies the Koszul formula, see Jeuris et al. (2012).

The parallel transport can be derived from the covariant derivative, and it follows that the parallel transport of a vector $w \in T_p(\mathcal{M})$ from a point $p \in \mathcal{M}$ along a geodesic curve in the direction of $v \in T_p(\mathcal{M})$ for time Δt is given by:

$$\mathfrak{T}(p, \Delta t v, w) \quad = \quad \mathrm{Exp}_p\left(\Delta t v/2\right) * p^{-1} * w. \tag{7.8}$$

Substituting $\Delta t v = \mathrm{Log}_p(q)$, we obtain the parallel transport $\Gamma_p^q : T_p(\mathcal{M}) \to T_q(\mathcal{M})$ that maps a vector in $T_p(\mathcal{M})$ to its parallel transported version along a geodesic curve in $T_q(\mathcal{M})$ given by:

$$\Gamma_p^q(w) \quad = \quad p^{1/2} * (p^{-1/2} * q)^{1/2} * p^{-1/2} * w. \tag{7.9}$$

Remark If $q = \mathrm{Id}$, where Id denotes the identity matrix, we obtain the so-called *whitening* transport as in e.g., Yuan et al. (2012), which parallel transports $w \in T_p(\mathcal{M})$ to $T_{\mathrm{Id}}(\mathcal{M})$ along a geodesic curve,

$$\Gamma_p^{\mathrm{Id}}(w) \quad = \quad p^{-1/2} * w \ \in T_{\mathrm{Id}}(\mathcal{M}). \tag{7.10}$$

Probability measures and random variables In order to perform statistics on the Riemannian manifold $(\mathcal{M}, g_R)$, we are concerned with the notions of probability distributions and random variables. A manifold-valued random variable $X : \Omega \to \mathcal{M}$ is a measurable function from some probability space $(\Omega, \mathcal{A}, \nu)$ to the measurable space $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$, where $\mathcal{B}(\mathcal{M})$ is the Borel algebra, i.e., the smallest σ-algebra containing all open sets in the complete separable metric space $(\mathcal{M}, \delta_R)$. In the following, we always work directly with the induced probability on $\mathcal{M}$, $\nu(B) = \nu(\{\omega \in \Omega : X(\omega) \in B\})$. By $P(\mathcal{M})$, we denote the set of all probability measures on $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$ and $P_p(\mathcal{M})$ denotes the subset of probability measures in $P(\mathcal{M})$ that have finite moments of order p with respect to the Riemannian distance δ_R, i.e., the L^p-Wasserstein space, see (Villani, 2009, Definition 6.4). That is,

$$P_p(\mathcal{M}) \quad := \quad \left\{ \nu \in P(\mathcal{M}) : \exists\, y_0 \in \mathcal{M},\ \text{s.t.} \int_{\mathcal{M}} \delta_R(y_0, x)^p \, \nu(dx) < \infty \right\}. \tag{7.11}$$

Note that if $\int_{\mathcal{M}} \delta_R(y_0, x)^p \, \nu(dx) < \infty$ for some $y_0 \in \mathcal{M}$ and $1 \leq p < \infty$, this is true for any $y \in \mathcal{M}$. This follows by the triangle inequality,

$$\int_{\mathcal{M}} \delta_R(y, x)^p \, \nu(dx) \ \leq \ 2^p \left(\delta_R(y, y_0)^p + \int_{\mathcal{M}} \delta_R(y_0, x)^p \, \nu(dx) \right) \ < \ \infty,$$

using that $\delta_R(p_1, p_2) < \infty$ for any $p_1, p_2 \in \mathcal{M}$ due to the Hopf-Rinow theorem for a geodesically complete manifold. For a sequence of probability measures $(\nu_n)_{n\in\mathbb{N}}$ in $P(\mathcal{M})$, $\nu_n \xrightarrow{w} \nu$ denotes weak convergence to the probability measure ν in the usual sense, i.e., $\int_{\mathcal{M}} \phi(x) \, \nu_n(dx) \to \int_{\mathcal{M}} \phi(x) \, \nu(dx)$ for every continuous and bounded function $\phi : \mathcal{M} \to \mathbb{R}$, and a sequence $(\nu_n)_{n\in\mathbb{N}}$ is said to be uniformly integrable if:

$$\lim_{K\to\infty} \sup_{n\in\mathbb{N}} \int_{\mathcal{M}} \delta_R(y_0, x) \mathbf{1}_{\{\delta_R(y_0,x) > K\}} \, \nu_n(dx) = 0, \quad \text{for some } y_0 \in \mathcal{M}.$$

Note that if $(\nu_n)_{n\in\mathbb{N}}$ is uniformly integrable for some $y_0 \in \mathcal{M}$, then the sequence is uniformly integrable for any $y \in \mathcal{M}$.

Intrinsic means Equipped with the notions of probability distributions and random variables on the manifold, we can characterize the center of a manifold-valued random variable X with probability measure ν. One important measure of centrality of a probability distribution ν on the manifold is the *intrinsic* mean, also *Karcher* or *Fréchet* mean, as its definition is intrinsic to the (Riemannian) distance on the space. The set of intrinsic means is given by the points that minimize the second moment with respect to the Riemannian distance δ_R,

$$\mu \;=\; \mathbb{E}_\nu[X] \;:=\; \arg\min_{y\in\text{supp}(\nu)} \int_{\mathcal{M}} \delta_R(y,x)^2 \; \nu(dx). \tag{7.12}$$

If $\nu \in P_2(\mathcal{M})$, then at least one Karcher mean exists as the above expectation is finite for each $y \in \mathcal{M}$. Moreover, since the manifold $(\mathcal{M}, g_R)$ is a geodesically complete manifold of non-positive curvature, (see Pennec et al. (2006) or Skovgaard (1984)), by (Le, 1995, Proposition 1) the Karcher mean μ is unique for any distribution $\nu \in P_2(\mathcal{M})$. By (Pennec, 2006, Corollary 1), the Karcher mean can also be represented by the unique point $\mu \in \mathcal{M}$ that satisfies,

$$\boldsymbol{E}_\nu[\text{Log}_\mu(X)] \;=\; \mathbf{0} \tag{7.13}$$

where $\mathbf{0}$ is the zero matrix and $\boldsymbol{E}_\nu[\cdot]$ is the Euclidean mean in the space of Hermitian matrices. In general, the sample intrinsic mean of a set of observations $\{X_1, \ldots, X_n\} \in \mathcal{M}$ has no closed-form solution, but it can be computed efficiently through a gradient descent algorithm as described in e.g., Pennec (2006).

Remark The representation of the intrinsic mean in eq.(7.13) above has an intuitive interpretation if we view the logarithmic map as a generalized notion of subtraction on the Riemannian manifold. In particular, if we equip the Riemannian manifold of HPD matrices with the Euclidean metric, (instead of the affine-invariant Riemannian metric), the logarithmic map reduces to ordinary matrix subtraction $\text{Log}_x(y) = y - x$ and the above representation becomes $\boldsymbol{E}_\nu[X - \mu] = \mathbf{0}$, or $\boldsymbol{E}_\nu[X] = \mu$.

8 Appendix II: Proofs

8.1 Proof of Proposition 3.1

Proof. Denote the distribution of $\mu_n := \mu_n(X_1, \ldots, X_n)$ by ν_n, we show recursively that:

$$\boldsymbol{E}[\delta_R(\mu_n, \mu)^2] \;=\; \int_{\mathcal{M}} \delta_R(x,\mu)\, d\nu_n(x) \;\leq\; \frac{1}{n}\boldsymbol{E}[\delta_R(X_1,\mu)^2].$$

By (Bhatia, 2009, Theorem 6.1.9), if $X_1, X_2, X_3 \in \mathcal{M}$, then for $t \in [0,1]$,

$$\delta_R(\eta(X_1, X_2, t), X_3)^2 \;\leq\; (1-t)\delta_R(X_1,X_3)^2 + t\delta_R(X_2,X_3)^2 - t(1-t)\delta_R(X_1,X_2)^2.$$

Substituting $X_3 = \mu$ and $t = 1/2$, (note that $\mu_2 = \eta(X_1, X_2, 1/2)$), and taking expectations on both sides yields:

$$\boldsymbol{E}_{X_1}\boldsymbol{E}_{X_2}[\delta_R(\mu_2,\mu)^2] \;\leq\; \frac{1}{2}\boldsymbol{E}_{X_1}[\delta_R(X_1,\mu)^2] + \frac{1}{2}\boldsymbol{E}_{X_2}[\delta_R(X_2,\mu)^2] - \frac{1}{4}\boldsymbol{E}_{X_1}\boldsymbol{E}_{X_2}[\delta_R(X_1,X_2)^2].$$

Using that $X_1, X_2 \overset{\text{iid}}{\sim} \nu$ we obtain,

$$\boldsymbol{E}[\delta_R(\mu_2,\mu)^2] \;\leq\; \boldsymbol{E}[\delta_R(X_1,\mu)^2] - \frac{1}{4}\boldsymbol{E}_{X_1}\boldsymbol{E}_{X_2}[\delta_R(X_1,X_2)^2]. \tag{8.1}$$

From the semi-parallelogram law above, (Ho et al., 2013, Proposition 1) derive:

$$\int_{\mathcal{M}} [\delta_R(x,y)^2 - \delta_R(x,\mu)^2]\, d\nu(x) \;\; \geq \;\; \delta_R(y,\mu)^2, \qquad \text{for any } y \in \mathcal{M}.$$

By the above inequality (and independence of X_1, X_2),

$$\begin{aligned} \boldsymbol{E}_{X_2}[\delta_R(X_1,X_2)^2 \mid X_1 = x_1] &= \int_{\mathcal{M}} \delta_R(x_1,X_2)^2\, d\nu(X_2) \\ &\geq \delta_R(x_1,\mu)^2 + \int_{\mathcal{M}} \delta_R(X_2,\mu)^2\, d\nu(X_2) \\ &= \delta_R(x_1,\mu)^2 + \boldsymbol{E}[\delta_R(X_2,\mu)^2], \end{aligned}$$

and consequently,

$$\begin{aligned} \boldsymbol{E}_{X_1}\boldsymbol{E}_{X_2}[\delta_R(X_1,X_2)^2] &\geq \int_{\mathcal{M}} \delta_R(X_1,\mu)^2\, d\nu(X_1) + \boldsymbol{E}[\delta_R(X_2,\mu)^2] \\ &= 2\boldsymbol{E}[\delta_R(X_1,\mu)^2]. \end{aligned}$$

Returning to eq.(8.1),

$$\boldsymbol{E}[\delta_R(\mu_2,\mu)^2] \;\leq\; \frac{1}{2}\boldsymbol{E}[\delta_R(X_1,\mu)^2].$$

Repeating the same argument, using independence of $\eta(X_1, X_2, 1/2)$ and $\eta(X_3, X_4, 1/2)$,

$$\boldsymbol{E}[\delta_R(\mu_4,\mu)^2] \;\leq\; \frac{1}{2}\boldsymbol{E}[\delta_R(\mu_2,\mu)^2] \;\leq\; \frac{1}{4}\boldsymbol{E}[\delta_R(X_1,\mu)^2].$$

Continuing this iteration up to μ_n, we find the upper bound:

$$\boldsymbol{E}[\delta_R(\mu_n,\mu)^2] \;\leq\; \frac{1}{2}\boldsymbol{E}_{n/2}[\delta_R(\mu_{n/2},\mu)^2] \;\leq\; \ldots \;\leq\; \frac{1}{n}\boldsymbol{E}[\delta_R(X_1,\mu)^2].$$

By Markov's inequality, $P(\delta_R(\mu_n,\mu) > \epsilon) \to 0$ for each $\epsilon > 0$ as $n \to \infty$, since the distribution of X_1 is assumed to have finite second moment with respect to δ_R, i.e., $\boldsymbol{E}[\delta_R(X_1,\mu)^2] < \infty$. $\square$

8.2 Proof of Proposition 3.2

Proof. Denote $L := (N-1)/2$, with $L \geq 0$, and fix $j \geq 1$ sufficiently large and $k \in [L, 2^{j-1} - (L+1)]$ away from the boundary, such that the neighboring $(j-1)$-midpoints $M_{j-1,k-L}, \ldots, M_{j-1,k+L}$ exist.

Remark: For $k < L$ or $k > 2^{j-1} - (L+1)$ near the boundary, we collect the N available closest neighbors of $M_{j-1,k}$ (either to the left or right). The remainder of the proof for the boundary case is exactly analogous to the non-boundary case and follows directly by mimicking the arguments outlined below.

We predict $M_{j,2k+1}$ from $M_{j-1,k-L}, \ldots, M_{j-1,k+L}$ via intrinsic polynomial interpolation of degree $N-1$ passing through the N points $\overline{M}_{j-1,0}, \ldots, \overline{M}_{j-1,N-1}$, where $\overline{M}_{j-1,k}$ denotes the cumulative intrinsic average as in eq.(2.4) in the main document. The predicted midpoint $\widetilde{M}_{j,2k+1}$ is then a weighted intrinsic average of the estimated polynomial at $(2k+1)2^{-j}$, i.e.,

$\widehat{M}_{(k-L)2^{-(j-1)}}((2k+1)2^{-j})$, and the given midpoint $\overline{M}_{j-1,L} = M_{(k-L)2^{-(j-1)}}(2k2^{-j})$, (with notation as in Section 2.1 in the main document).

For notational simplicity, write $M(t) := M_{(k-L)2^{-(j-1)}}(t)$ and $\widehat{M}(t) := \widehat{M}_{(k-L)2^{-(j-1)}}(t)$ for the true and estimated intrinsic cumulative mean curves respectively, where the latter is an interpolating polynomial of order $N-1$ passing through N equidistant points $x_0, \ldots, x_{N-1}$ on the curve $M(t)$. $M(t)$ itself is a smooth curve with existing covariant derivatives up to order N, and $|x_0 - x_{N-1}| \lesssim 2^{-j}$. The polynomial remainder of the interpolating polynomial in Newton form with respect to the smooth curve, for every $x \in [(k-L)2^{-(j-1)}, (k+L)2^{-(j-1)}]$, is upper bounded by:

$$\frac{d}{dt}\widehat{M}(t)|_{t=x} - \frac{d}{dt}M(t)|_{t=x} \;\lesssim\; \frac{(x-x_0)\cdots(x-x_{N-1})}{N!}\,\Gamma(M)^x_\xi\left(\nabla^N_{\frac{d}{dt}M}\frac{d}{dt}M\big|_{t=\xi}\right) \;=\; O(2^{-jN})$$

for some $\xi \in [(k-L)2^{-(j-1)}, (k+L)2^{-(j-1)}]$ by the mean value theorem for divided differences. This is closely related to the Taylor expansion in eq.(3.2) in the main document. In particular, the limit of the Newton polynomial if all nodes coincide is the Taylor polynomial, as the divided differences become covariant derivatives, and the covariant derivatives in the Taylor expansions of the Taylor polynomial and the smooth curves match up to order $N-1$.

By definition of the derivative $\widehat{M}'(t) := \frac{d}{dt}\widehat{M}(t) = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\mathrm{Log}_{\widehat{M}(t)}(\widehat{M}(t+\Delta t))$ and the fundamental theorem of calculus, it is verified that:

$$\widehat{M}(t+\Delta t) \;=\; \mathrm{Exp}_{\widehat{M}(t)}\left(\int_t^{t+\Delta t}\widehat{M}'(u)\,du\right).$$

Substituting $t = 2k2^{-j}$ and $\Delta t = 2^{-j}$ and using that $\widehat{M}(2k2^{-j}) = M(2k2^{j})$ by construction, we obtain:

$$\begin{aligned}\widehat{M}((2k+1)2^{-j}) &= \mathrm{Exp}_{M(2k2^{-j})}\left(\int_{2k2^{-j}}^{(2k+1)2^{-j}}\widehat{M}'(u)\,du\right)\\ &= \mathrm{Exp}_{M(2k2^{-j})}\left(\int_{2k2^{-j}}^{(2k+1)2^{-j}}M'(u)\,du + O(2^{-jN})\right). \qquad (8.2)\end{aligned}$$

The second step in the above equation follows immediately if $L=0$ (i.e., $N=1$), since,

$$\int_{2k2^{-j}}^{(2k+1)2^{-j}}\widehat{M}'(u)\,du \;=\; \int_{2k2^{-j}}^{(2k+1)2^{-j}}[M'(u)+O(1)]\,du \;=\; \int_{2k2^{-j}}^{(2k+1)2^{-j}}M'(u)\,du + O(2^{-j}).$$

If $L \geq 1$, the second step in eq.(8.2) follows by the polynomial remainder error bound above, since $\widehat{M}'(u) = M'(u) + O(2^{-jN})$ for each $u \in [2k2^{-j}, (2k+1)2^{-j}] \subset [(k-L)2^{-(j-1)}, (k+L)2^{-(j-1)}]$.

Application of the logarithmic map $\mathrm{Log}_{M(2k2^{-j})}(\cdot)$ to both sides in eq.(8.2) and using that $\mathrm{Log}_{M(t)}(M(t+\Delta t)) = \int_t^{t+\Delta t}M'(u)\,du$ as above, we rewrite:

$$\mathrm{Log}_{M(2k2^{-j})}(\widehat{M}((2k+1)2^{-j}) \;=\; \mathrm{Log}_{M(2k2^{-j})}(M(2k+1)2^{-j}) + O(2^{-jN}). \qquad (8.3)$$

For notational convenience, in the remainder of this proof, we write $\Lambda = \lambda E$ for some arbitrary (not necessarily fixed) deterministic matrix $E \in \mathbb{C}^{d\times d}$ and constant $\lambda \lesssim 2^{-jN}$, i.e., $\Lambda =$

$O(2^{-jN})$.

Let $M, M_1, M_2 \in \mathcal{M}$ be deterministic matrices, we verify the following implication:

Claim. *If* $\text{Log}_M(M_1) - \text{Log}_M(M_2) = O(\lambda)$*, then also* $M_1 = M_2 + O(\lambda)$.

Proof. Starting from $\text{Log}_M(M_1) - \text{Log}_M(M_2) = O(\lambda)$, by the definition of the logarithmic map, we write out,

$$\begin{aligned} M^{1/2} * \text{Log}(M^{-1/2} * M_1) &= M^{1/2} * \text{Log}(M^{-1/2} * M_2) + O(\lambda) \quad \Rightarrow \\ \text{Log}(M^{-1/2} * M_1) &= \text{Log}(M^{-1/2} * M_2) + O(\lambda) \quad \Rightarrow \\ M^{-1/2} * M_1 &= \text{Exp}(\text{Log}(M^{-1/2} * M_2) + O(\lambda)). \end{aligned}$$

For $\lambda \to 0$ sufficiently small, $M_1 = \text{Exp}(\text{Log}(M_2) + O(\lambda))$ also implies $M_1 = M_2 + O(\lambda)$. This follows by Taylor expanding the matrix exponential,

$$\begin{aligned} M_1 &= \text{Exp}(\text{Log}(M_2) + O(\lambda)) = \sum_{k=0}^{\infty} \frac{(\text{Log}(M_2) + O(\lambda))^k}{k!} \\ &= \sum_{k=0}^{\infty} \frac{(\text{Log}(M_2))^k + O(\lambda)}{k!} = \sum_{k=0}^{\infty} \frac{(\text{Log}(M_2))^k}{k!} + O(\lambda) \sum_{k=0}^{\infty} \frac{1}{k!} = M_2 + O(\lambda). \end{aligned}$$

As a consequence, also,

$$\begin{aligned} M^{-1/2} * M_1 &= \text{Exp}(\text{Log}(M^{-1/2} * M_2) + O(\lambda)) \quad \Rightarrow \\ M^{-1/2} * M_1 &= M^{-1/2} * M_2 + O(\lambda) \quad \Rightarrow \\ M^{-1/2} * (M_1 - M_2) &= O(\lambda) \quad \Rightarrow \\ M_1 &= M_2 + O(\lambda). \end{aligned}$$

□

Applying the above implication to eq.(8.3) yields,

$$\widehat{M}((2k+1)2^{-j}) = M((2k+1)2^{-j}) + O(2^{-jN}). \tag{8.4}$$

The predicted midpoint $\widetilde{M}_{j,2k+1}$ is reconstructed from $\widehat{M}((2k+1)2^{-j})$ and $M(2k2^{-j})$ as follows. By definition of $M(t)$ as the cumulative intrinsic mean curve, we can write $M((2k+1)2^{-j})$ as a weighted intrinsic average between $\overline{M}_{j-1,L} = M(2k2^{-j})$ and $M_{j,2k+1}$ according to:

$$\begin{aligned} M((2k+1)2^{-j}) = \text{Exp}_{M((2k+1)2^{-j})}\Bigg(&\frac{(N-1)2^{-j}}{N2^{-j}} \text{Log}_{M((2k+1)2^{-j})}(\overline{M}_{j-1,L}) \\ &+ \frac{2^{-j}}{N2^{-j}} \text{Log}_{M((2k+1)2^{-j})}(M_{j,2k+1}) \Bigg). \end{aligned}$$

Application of the logarithmic map $\text{Log}_{M((2k+1)2^{-j})}(\cdot)$ to both sides and rearranging terms (substitute $N - 1 = 2L$), gives,

$$\frac{-2L}{N} \text{Log}_{M((2k+1)2^{-j})}(\overline{M}_{j-1,L}) = \frac{1}{N} \text{Log}_{M((2k+1)2^{-j})}(M_{j,2k+1}).$$

Or in terms of $M_{j,2k+1}$,

$$\begin{aligned} M_{j,2k+1} &= \mathrm{Exp}_{M((2k+1)2^{-j})}\left(-2L \cdot \mathrm{Log}_{M((2k+1)2^{-j})}(\overline{M}_{j-1,L})\right) \\ &= \eta\left(M((2k+1)2^{-j}), \overline{M}_{j-1,L}, -2L\right). \end{aligned}$$

The predicted midpoint $\widetilde{M}_{j,2k+1}$ is given by replacing the true point $M((2k+1)2^{-j})$ by the estimated point $\widehat{M}((2k+1)2^{-j})$, ($\overline{M}_{j-1,L}$ is known), i.e.,

$$\widetilde{M}_{j,2k+1} = \eta\left(\widehat{M}((2k+1)2^{-j}), \overline{M}_{j-1,L}, -2L\right). \tag{8.5}$$

Below, we use that $(M+\Lambda)^a = M^a + O(\lambda)$ for $a \in \mathbb{N}$, $(M+\Lambda)^{1/2} = M^{1/2} + O(\lambda)$ and $(M+\Lambda)^{-1} = M^{-1} + O(\lambda)$ for $M \in \mathcal{M}$ and $\lambda \to 0$ sufficiently small, as verified in the proof of Proposition 3.3, (note that this is the deterministic version), combined with eq.(8.4) and the definition of the geodesic in eq.(7.3). Writing out eq.(8.5) gives,

$$\begin{aligned} \widetilde{M}_{j,2k+1} &= \left(M((2k+1)2^{-j})^{1/2} + \Lambda\right) * \left(\left(M((2k+1)2^{-j})^{-1/2} + \Lambda\right) * \overline{M}_{j-1,L}\right)^{-2L} \\ &= \left(M((2k+1)2^{-j})^{1/2} + \Lambda\right) * \left(\left(M((2k+1)2^{-j})^{-1/2} * \overline{M}_{j-1,L}\right)^{-1} + \Lambda\right)^{2L} \\ &= \left(M((2k+1)2^{-j})^{1/2} + \Lambda\right) * \left(\left(M((2k+1)2^{-j})^{-1/2} * \overline{M}_{j-1,L}\right)^{-2L} + \Lambda\right) \\ &= M_{j,2k+1} + O(2^{-jN}). \end{aligned} \tag{8.6}$$

Substituting the above result in the whitened wavelet coefficient $\mathfrak{D}_{j,k} = 2^{-j/2}\mathrm{Log}(\widetilde{M}_{j,2k+1}^{-1/2} * M_{j,2k+1})$, by the same identities as used above combined with $\mathrm{Log}(M+\Lambda) = \mathrm{Log}(M) + O(\lambda)$, (verified in the proof of Proposition 3.3), it follows that for $j \geq 1$ sufficiently large,

$$\begin{aligned} \|\mathfrak{D}_{j,k}\|_F &= \left\|2^{-j/2}\mathrm{Log}\big((M_{j,2k+1}+\Lambda)^{-1/2} * M_{j,2k+1}\big)\right\|_F \\ &= 2^{-j/2}\left\|\mathrm{Log}\big((M_{j,2k+1}^{-1/2}+\Lambda) * M_{j,2k+1}\big)\right\|_F \\ &= 2^{-j/2}\left\|\mathrm{Log}\big(\mathrm{Id}+\Lambda\big)\right\|_F = O\left(2^{-j/2}2^{-jN}\right), \end{aligned}$$

where in the final step we expanded $\mathrm{Log}(\mathrm{Id}+\Lambda) = O(2^{-jN})$ via its Mercator series (see (Higham, 2008, Section 11.3)), using that the spectral radius of Λ is smaller than 1 for j sufficiently large. □

8.3 Proof of Proposition 3.3

Proof. By the proof of Proposition 3.1, $\boldsymbol{E}[\delta_R(M_{j,k,n}, M_{j,k})^2] = O(2^{-(J-j)})$ for each $j \geq 0$ and $0 \leq k \leq 2^j - 1$. For notational convenience, in the remainder of this proof $\epsilon_{j,n}$ denotes a general (not necessarily the same) random error matrix that satisfies $\boldsymbol{E}\|\epsilon_{j,n}\|_F^2 = O(2^{-(J-j)})$. Furthermore, we can appropriately write $M_{j,k,n} = \mathrm{Exp}_{M_{j,k}}(\epsilon_{j,n})$, such that $M_{j,k,n} \xrightarrow{p} M_{j,k}$ as $J \to \infty$ at the correct rate since,

$$\begin{aligned} \boldsymbol{E}[\delta_R(\mathrm{Exp}_{M_{j,k}}(\epsilon_{j,n}), M_{j,k})^2] &= \boldsymbol{E}\|\mathrm{Log}(M_{j,k}^{-1/2} * \mathrm{Exp}_{M_{j,k}}(\epsilon_{j,n}))\|_F^2 \\ &= \boldsymbol{E}\|M_{j,k}^{-1/2} * \epsilon_{j,n}\|_F^2, \\ &= O(2^{-(J-j)}) \end{aligned}$$

using the definitions of the Riemannian distance function and the logarithmic and exponential maps. In particular, by a first-order Taylor expansion of the matrix exponential, (abusing notation of $\epsilon_{j-1,n}$), $M_{j-1,k,n} = M_{j-1,k}^{1/2} * \mathrm{Exp}(\epsilon_{j-1,n}) = M_{j-1,k}^{1/2} * (\mathrm{Id} + \epsilon_{j-1,n} + \ldots) = M_{j-1,k} + \epsilon_{j-1,n}$.

By eq.(2.5) in the main document, the predicted midpoint $\widetilde{M}_{j,2k+1,n}$ is a weighted intrinsic mean of N coarse-scale midpoints $(M_{j-1,k,n})_k$ with weights summing up to 1. The rate of $\widetilde{M}_{j,2k+1,n}$ is therefore upper bounded by the (worst) convergence rate of the individual midpoints $(M_{j-1,k,n})_k$, and we can also write $\widetilde{M}_{j,2k+1,n} = \widetilde{M}_{j,2k+1} + \epsilon_{j-1,n}$.

Below, we verify several implications that are needed to finish the proof. let $M \in \mathcal{M}$ be a deterministic matrix and $\lambda E = O_p(\lambda)$ a random error matrix, such that $\boldsymbol{E}\|\lambda E\|_F = O(\lambda)$.

Claim. If $\lambda \to 0$ sufficiently small, then $\mathrm{Log}(M + \lambda E) \;=\; \mathrm{Log}(M) + O_p(\lambda)$.

Proof. Rewrite $\mathrm{Log}(M + \lambda E) = \mathrm{Log}(M(\mathrm{Id} + \lambda M^{-1} E))$. By the Baker-Campbell-Hausdorff formula (e.g., (Higham, 2008, Theorem 10.4)), with $X = \mathrm{Log}(M)$ and $Y = \mathrm{Log}(\mathrm{Id} + \lambda M^{-1} E))$,

$$\mathrm{Log}(M + \lambda E) \;=\; X + Y + \frac{1}{2}[X, Y] + \frac{1}{12}([X, [X, Y]] + [Y, [Y, X]]) + \frac{1}{24}[Y, [X, [X, Y]]] - \ldots,$$

where $[X, Y] = XY - YX$ denotes the commutator of X and Y. In particular,

$$\begin{aligned}
[X, Y] &= [\mathrm{Log}(M), \mathrm{Log}(\mathrm{Id} + \lambda M^{-1} E)] \\
&= \mathrm{Log}(M)\mathrm{Log}(\mathrm{Id} + \lambda M^{-1} E) - \mathrm{Log}(\mathrm{Id} + \lambda M^{-1} E)\mathrm{Log}(M) \\
&= \mathrm{Log}(M)(\lambda M^{-1} E + O_p(\lambda^2)) - (\lambda M^{-1} E + O_p(\lambda^2))\mathrm{Log}(M) \\
&= O_p(\lambda).
\end{aligned}$$

Here, we expanded $\mathrm{Log}(\mathrm{Id} + \lambda M^{-1} E) = \lambda M^{-1} E + O_p(\lambda^2)$ via its Mercator series (e.g., (Higham, 2008, Section 11.3)), using that the spectral radius $\rho(\lambda M^{-1} E) = \lambda \rho(M^{-1} E) < 1$ almost surely for $\lambda \to 0$ sufficiently small.

Iterating the above argument, it follows that all the nested (higher-order) commutators are of the order $O_p(\lambda)$ as well, and we rewrite:

$$\mathrm{Log}(M + \lambda E) \;=\; \mathrm{Log}(M) + \mathrm{Log}(\mathrm{Id} + \lambda M^{-1} E) + O_p(\lambda).$$

Expanding again $\mathrm{Log}(\mathrm{Id} + \lambda M^{-1} E) = \lambda M^{-1} E + O_p(\lambda^2) = O_p(\lambda)$, (for λ sufficiently small), the claim follows. □

Claim. If $\lambda \to 0$ sufficiently small, then $(M + \lambda E)^{1/2} \;=\; M^{1/2} + O_p(\lambda)$ and $(M + \lambda E)^{-1} \;=\; M^{-1} + O_p(\lambda)$.

Proof. For the first claim, Taylor expanding the matrix exponential,

$$\begin{aligned}
(M + \lambda E)^{1/2} &= \mathrm{Exp}\left(\frac{1}{2}\mathrm{Log}(M + \lambda E)\right) \;=\; \sum_{k=0}^{\infty} \frac{(\mathrm{Log}(M + \lambda E))^k}{2^k k!} \\
&= \sum_{k=0}^{\infty} \frac{(\mathrm{Log}(M) + O_p(\lambda))^k}{2^k k!} \;=\; \sum_{k=0}^{\infty} \frac{(\mathrm{Log}(M))^k}{2^k k!} + O_p(\lambda) \;=\; M^{1/2} + O_p(\lambda),
\end{aligned}$$

using the previous claim $\mathrm{Log}(M+\lambda E) = \mathrm{Log}(M) + O_p(\lambda)$ for $\lambda \to 0$ sufficiently small.

For the second claim, rewrite, (for λ sufficiently small),

$$\begin{aligned}(M+\lambda E)^{-1} &= (M(\mathrm{Id}+\lambda M^{-1}E))^{-1} \\ &= (\mathrm{Id}+\lambda M^{-1}E))^{-1}M^{-1} \\ &= (\mathrm{Id}-\lambda M^{-1}E+(\lambda M^{-1}E)^2-\ldots)M^{-1} \;=\; M^{-1}+O_p(\lambda),\end{aligned}$$

applying a binomial series expansion of the matrix inverse $(\mathrm{Id}+\lambda M^{-1}E))^{-1}$, using that the spectral radius $\rho(\lambda M^{-1}E) = \lambda\rho(M^{-1}E) < 1$ almost surely for $\lambda \to 0$ sufficiently small. Combining the two claims, we find in particular also that $(M+\lambda E)^{-1/2} = M^{-1/2} + O_p(\lambda)$. □

Combining the above results, for $j < J$ sufficiently small such that the above claims hold, we write out for the empirical whitened wavelet coefficient $\widehat{\mathfrak{D}}_{j,k,n}$, (with some abuse of notation for $\epsilon_{j,n}$),

$$\begin{aligned}\widehat{\mathfrak{D}}_{j,k,n} &= 2^{-j/2}\,\mathrm{Log}\left((\widetilde{M}_{j,2k+1}+\epsilon_{j-1,n})^{-1/2} * (M_{j,2k+1}+\epsilon_{j,n})\right) \\ &= 2^{-j/2}\,\mathrm{Log}\left((\widetilde{M}^{-1/2}_{j,2k+1}+\epsilon_{j-1,n}) * (M_{j,2k+1}+\epsilon_{j,n})\right) \\ &= 2^{-j/2}\,\mathrm{Log}\left(\widetilde{M}^{-1/2}_{j,2k+1} * M_{j,2k+1} + \epsilon_{j,n} + \ldots\right) \\ &= 2^{-j/2}\,\mathrm{Log}\left(\widetilde{M}^{-1/2}_{j,2k+1} * M_{j,2k+1}\right) + 2^{-j/2}\,O_p(2^{-(J-j)/2}) \\ &= \mathfrak{D}_{j,k} + 2^{-j/2}\,O_p(2^{-(J-j)/2}).\end{aligned}$$

Plugging in the above result, it follows that for $j < J$ sufficiently small,

$$\boldsymbol{E}\|\widehat{\mathfrak{D}}_{j,k,n} - \mathfrak{D}_{j,k}\|_F^2 \;=\; O(2^{-j}\,2^{-(J-j)}) \;=\; O(n^{-1}).$$

□

8.4 Proof of Theorem 3.4

Proof. For the first part of the theorem, suppose that $J_0 = \log_2(n)/(2N+1) \gg 1$ is sufficiently large such that the rates in Propositions 3.2 and 3.3 hold. Then,

$$\begin{aligned}\sum_{j,k}\boldsymbol{E}\|\widehat{\mathfrak{D}}_{j,k} - \mathfrak{D}_{j,k}\|_F^2 &= \sum_{j\geq J_0}\|\mathfrak{D}_{j,k}\|_F^2 + \sum_{j<J_0}\boldsymbol{E}\|\widehat{\mathfrak{D}}_{j,k} - \mathfrak{D}_{j,k}\|_F^2 \\ &\lesssim \sum_{j\geq J_0} 2^j(2^{-j}2^{-2jN}) + \sum_{j<J_0} 2^j n^{-1} \\ &= \left(\sum_{j=0}^{J}(2^{-2N})^j - \sum_{j=0}^{J_0-1}(2^{-2N})^j\right) + n^{-1}\sum_{j=1}^{J_0-1}2^j \\ &= \frac{(2^{-2N})^{J_0} - (2^{-2N})^{(J+1)}}{1-2^{-2N}} + n^{-1}(2^{J_0}-2) \\ &\lesssim (2^{-2N})^{J_0} + n^{-1}2^{J_0} + n^{-1} \\ &\lesssim n^{-2N/(2N+1)}, \qquad (8.7)\end{aligned}$$

where the last step follows from substituting $J_0 = \log_2(n)/(2N+1)$ since,

$$\begin{array}{rclcrcl} (2^{-2N})^{J_0} & = & \exp(-2NJ_0\log(2)) & = & \exp\left(\frac{-2N}{2N+1}\log(n)\right) & = & n^{-2N/(2N+1)} \\ n^{-1}2^{J_0} & = & \exp(-\log(n)+J_0\log(2)) & = & \exp\left(\frac{-2N}{2N+1}\log(n)\right) & = & n^{-2N/(2N+1)}. \end{array}$$

For the second part of the theorem, if we can verify that $\boldsymbol{E}[\delta_R(M_{J,k}, \widehat{M}_{J,k,n})^2] \lesssim n^{-2N/(2N+1)}$ for each $k = 0, \ldots, n-1$, the proof is finished.

At scales $j = 1, \ldots, J$, based on the estimated midpoints $(\widehat{M}_{j-1,k',n})_{k'}$ and the estimated wavelet coefficient $\widehat{D}_{j,k,n}$, in the inverse wavelet transform, the finer-scale midpoint $\widehat{M}_{j,k,n}$ is estimated through,

$$\widehat{M}_{j,k,n} \;=\; \mathrm{Exp}_{\widehat{\widetilde{M}}_{j,k,n}}\left(2^{j/2}\widehat{D}_{j,k,n}\right).$$

where $\widehat{\widetilde{M}}_{j,k,n}$ is the predicted midpoint at scale-location (j,k) based on $(\widehat{M}_{j-1,k',n})_{k'}$. In particular, at scale $j = 1$, $\widehat{\widetilde{M}}_{1,k,n} = \widetilde{M}_{1,k,n}$ as the estimated coarsest midpoints $(\widehat{M}_{0,k',n})_{k'}$ correspond to the empirical coarsest midpoints $(M_{0,k',n})_{k'}$.

At scales $j = 1, \ldots, J_0 - 1$, we do not alter the wavelet coefficients. Assuming that $j \ll J$ is sufficiently small, such that the rate in Proposition 3.3 holds, we write $\widehat{\mathfrak{D}}_{j,k,n} = \mathfrak{D}_{j,k} + \eta_n$, with η_n a general (not always the same) random error matrix satisfying $\boldsymbol{E}\|\eta_n\|_F = O(n^{-1/2})$. Also, by the proof of Proposition 3.3 (using the same notation), we can write $\widetilde{M}_{j,k,n} = \widetilde{M}_{j,k} + \epsilon_{j,n}$, where $\epsilon_{j,n}$ is a general (not always the same) random error matrix satisfying $\boldsymbol{E}\|\epsilon_{j,n}\|_F = O(2^{-(J-j)/2})$.

In particular, at scale $j = 1$,

$$\begin{array}{rcl} \widehat{M}_{1,k,n} & = & \mathrm{Exp}_{\widehat{\widetilde{M}}_{1,k,n}}\left(2^{1/2}\widehat{D}_{1,k,n}\right) \\ & = & \widetilde{M}_{1,k,n}^{1/2} * \mathrm{Exp}\left(2^{1/2}\widetilde{M}_{1,k,n}^{-1/2} * \widehat{D}_{1,k,n}\right) \\ & = & \widetilde{M}_{1,k,n}^{1/2} * \mathrm{Exp}\left(2^{1/2}\widehat{\mathfrak{D}}_{1,k,n}\right) \\ & = & \left(\widetilde{M}_{1,k} + \epsilon_{1,n}\right)^{1/2} * \mathrm{Exp}\left(2^{1/2}(\mathfrak{D}_{1,k} + \eta_n)\right) \\ & = & \left(\widetilde{M}_{1,k}^{1/2} + \epsilon_{1,n}\right) * \left(\mathrm{Exp}(2^{1/2}\mathfrak{D}_{1,k}) + 2^{1/2}\eta_n\right) \\ & = & M_{1,k} + O_p(2^{1/2}n^{-1/2}) + O_p(2^{-(J-1)/2}) \\ & = & M_{1,k} + O_p(2^{1/2}n^{-1/2}). \end{array} \tag{8.8}$$

Here, we used that $(M + \lambda E)^{1/2} = M^{1/2} + O_p(\lambda)$ for $\lambda \to 0$ sufficiently small as in the proof of Proposition 3.3, and a Taylor expansion of the matrix exponential:

$$\mathrm{Exp}(D + \eta_n) \;=\; \sum_{k=0}^{\infty} \frac{(D+\eta_n)^k}{k!} \;=\; \sum_{k=0}^{\infty} \frac{D^k}{k!} + O_p(n^{-1/2}) \;=\; \mathrm{Exp}(D) + O_p(n^{-1/2}).$$

Iterating this same argument for each scale $j = 2, \ldots, J_0 - 1$, we find that:

$$\widehat{M}_{J_0-1,k,n} \;=\; M_{J_0-1,k} + \sum_{j=1}^{J_0-1} O_p(n^{-1/2}2^{j/2}) \;=\; M_{J_0-1,k} + O_p(n^{-1/2}2^{(J_0-1)/2}).$$

As a consequence, (as in the proof of Proposition 3.3), we can write $\widehat{\widetilde{M}}_{J_0,k,n} = \widetilde{M}_{J_0,k} + \epsilon_{J_0,n}$, where $\epsilon_{J_0,n} = O_p(n^{-1/2}2^{J_0/2})$. At scales $j = J_0, \ldots, J$, we set $\widehat{D}_{j,k,n} = \mathbf{0}$ for each k. Assuming that $j \gg 1$ is sufficiently large, such that the rate in Proposition 3.2 holds, we can write $\widehat{D}_{j,k,n} = \mathbf{0} = \mathfrak{D}_{j,k} + \zeta_{j,N}$, with $\zeta_{j,N}$ a general (not always the same) deterministic error matrix satisfying $\|\zeta_{j,N}\|_F = O(2^{-j/2}2^{-jN})$.

In particular, at scale $j = J_0$,

$$\begin{aligned}
\widehat{M}_{J_0,k,n} &= \mathrm{Exp}_{\widehat{\widetilde{M}}_{J_0,k,n}}\big(2^{J_0/2}\widehat{D}_{J_0,k,n}\big) \\
&= \big(\widetilde{M}_{J_0,k} + \epsilon_{J_0,n}\big)^{1/2} * \mathrm{Exp}\Big(\big(\widetilde{M}_{J_0,k} + \epsilon_{J_0,n}\big)^{-1/2} * 2^{J_0/2}\big(\mathfrak{D}_{J_0,k} + \zeta_{J_0,n}\big)\Big) \\
&= \big(\widetilde{M}^{1/2}_{J_0,k} + \epsilon_{J_0,n}\big) * \mathrm{Exp}\Big(\big(\widetilde{M}^{-1/2}_{J_0,k} + \epsilon_{J_0,n}\big) * \big(2^{J_0/2}\mathfrak{D}_{J_0,k} + 2^{J_0/2}\zeta_{J_0,n}\big)\Big) \\
&= \Big(\widetilde{M}^{1/2}_{J_0,k} + \epsilon_{J_0,n}\Big) * \Big(\mathrm{Exp}(2^{J_0/2}D_{J_0,k}) + 2^{J_0/2}\epsilon_{J_0,n}\mathfrak{D}_{J_0,k} + 2^{J_0/2}\zeta_{J_0,n}\Big) \\
&= \Big(\widetilde{M}^{1/2}_{J_0,k} + \epsilon_{J_0,n}\Big) * \Big(\mathrm{Exp}(2^{J_0/2}D_{J_0,k}) + O_p\big(2^{-J_0N}\big)\Big) \\
&= M_{J_0,k} + O_p(n^{-1/2}2^{J_0/2}) + O_p\big(2^{-J_0N}\big),
\end{aligned}$$

which follows in the same way as in eq.(8.8) above, combined with the observation that $2^{J_0/2}\epsilon_{J_0,n}\mathfrak{D}_{J_0,k} = O_p(2^{-J_0N})$, since $\|2^{J_0/2}\epsilon_{J_0,n}\mathfrak{D}_{J_0,k}\|_F = O_p(2^{-(J-J_0)/2}2^{-J_0N}) = O_p(2^{-J_0N})$ by Proposition 3.2. Iterating this same argument for each scale $j = J_0 + 1, \ldots, J$ yields,

$$\widehat{M}_{J,k,n} \;=\; M_{J,k} + O_p(n^{-1/2}2^{J_0/2}) + \sum_{j=J_0}^{J} O_p\big(2^{-jN}\big) \;=\; M_{J,k} + O_p\big(2^{-J_0N}\big) + O_p\big(n^{-1/2}2^{J_0/2}\big).$$

Plugging in $J_0 = \log_2(n)/(2N+1)$, as previously demonstrated, the above expression reduces to:

$$\widehat{M}_{J,k,n} \;=\; M_{J,k} + O_p\big(n^{-N/(2N+1)}\big), \qquad \text{for each } k = 0, \ldots, n-1.$$

For notational convenience, denote by $\xi_{n,N}$ a general (not always the same) random error matrix such that $\boldsymbol{E}\|\xi_{n,N}\|_F = O(n^{-N/(2N+1)})$. For each $k = 0, \ldots, n-1$, by the previous result:

$$\begin{aligned}
\boldsymbol{E}\Big[\delta_R(M_{J,k}, \widehat{M}_{J,k,n})^2\Big] &= \boldsymbol{E}\Big[\delta_R\big(M_{J,k}, M_{J,k} + \xi_{n,N}\big)^2\Big] \\
&= \boldsymbol{E}\Big\|\mathrm{Log}\Big(M_{J,k}^{-1/2} * \big(M_{J,k} + \xi_{n,N}\big)\Big)\Big\|_F^2 \\
&= \boldsymbol{E}\,\big\|\mathrm{Log}\big(\mathrm{Id} + \xi_{n,N}\big)\big\|_F^2 \;=\; O(n^{-2N/(2N+1)}),
\end{aligned}$$

where in the final step we expanded $\mathrm{Log}(\mathrm{Id} + \xi_{n,N}) = O_p(n^{-N/(2N+1)})$ via its Mercator series, using that the spectral radius of $\xi_{n,N}$ is smaller than 1 almost surely for n sufficiently large. □

8.4.1 Proof of remark Theorem 3.4

Let $\gamma_n(t) = \gamma(t) + \epsilon_{n,N}$ and $\hat{\gamma}(t)$ be as defined in the remark after Theorem 3.4, with $\epsilon_{n,N}$ a general error matrix, such that $\|\epsilon_{n,N}\|_F = O(n^{-N/(2N+1)})$. Then we can upper bound,

$$\begin{aligned}
\delta(\gamma(t), \gamma_n(t))^2 &= \|\mathrm{Log}\big(\gamma(t)^{-1/2} * (\gamma(t) + \epsilon_n)\big)\|_F^2 \\
&= \|\mathrm{Log}(\mathrm{Id} + \epsilon_{n,N})\|_F^2 \;=\; O(n^{-2N/(2N+1)}),
\end{aligned}$$

where in the final step we again expand $\mathrm{Log}(\mathrm{Id} + \epsilon_{n,N}) = O(n^{-N/(2N+1)})$ via its Mercator series, provided that n is sufficiently large.

By the triangle inequality, the integrated mean-squared error of the linear wavelet estimator with respect to the continuous curve γ then also satisfies,

$$\begin{aligned}\int_0^1 \boldsymbol{E}\left[\delta_R(\hat{\gamma}_n(t),\gamma(t))^2\right]\,dt &\leq 2^2\left(\int_0^1 \boldsymbol{E}\left[\delta_R(\hat{\gamma}_n(t),\gamma_n(t))^2\right]\,dt + \int_0^1 \delta_R(\gamma_n(t),\gamma(t))^2\,dt\right)\\ &= 2^2\left(\frac{1}{n}\sum_{k=0}^{n-1}\boldsymbol{E}\left[\delta_R(\widehat{M}_{J,k,n},M_{J,k})^2\right] + \int_0^1 \delta_R(\gamma_n(t),\gamma(t))^2\,dt\right)\\ &\lesssim n^{-2N/(2N+1)},\end{aligned}$$

using the convergence rate for the linear wavelet estimator derived above.

8.5 Proof of Theorem 4.1

Proof. First, we derive the bias $b(X,f) = c(d,L)\cdot f$. By linearity of the (ordinary) expectation:

$$b(X,f) \;=\; \boldsymbol{E}[\mathrm{Log}_f(X)] \;=\; f^{1/2} * \boldsymbol{E}[\mathrm{Log}(f^{-1/2} * X)], \tag{8.9}$$

using that $g * \mathrm{Log}_{X_1}(X_2) = \mathrm{Log}_{g*X_1}(g * X_2)$ for any $g \in \mathrm{GL}(d,\mathbb{C})$. The transformed random variable $Y := f^{-1/2} * X$ is distributed as $Y \sim W_d^c(L, L^{-1}\mathrm{Id})$, which is unitarily invariant (see e.g., (Muirhead, 1982, Section 3.2)). By (Tulino and Verdú, 2004, Section 2.1.5), taking the eigendecomposition of a unitarily invariant matrix $Y = Q * \Lambda$, the matrix of eigenvectors Q is distributed according to the Haar measure, i.e., the uniform distribution on the set of unitary matrices $\mathcal{U}_d = \{U \in \mathrm{GL}(d,\mathbb{C}) \mid U^*U = \mathrm{Id}\}$, implying that the eigenvectors $(\vec{q}_i)_{i=1,\ldots,d}$ (the columns of Q) are identically distributed. Furthermore, Q is independent of the diagonal eigenvalue-matrix Λ, therefore (see also Smith (2000)):

$$\boldsymbol{E}[\mathrm{Log}(Y)] \;=\; \boldsymbol{E}\left[\sum_{i=1}^{d}\log(\lambda_i)\vec{q}_i\vec{q}_i^{\,*}\right] \;=\; \boldsymbol{E}[\vec{q}_i\vec{q}_i^{\,*}]\boldsymbol{E}[\log(\det(\Lambda))]. \tag{8.10}$$

Since Y is Hermitian, $Q \in \mathcal{U}_d$, and therefore $\boldsymbol{E}[\log(\det(\Lambda))] \;=\; \boldsymbol{E}[\log(\det(Y))]$. By (Muirhead, 1982, Theorem 3.2.15),

$$\log(\det(Y)) \;\sim\; -d\log(2L) + \sum_{i=1}^{d}\log\left(\chi^2_{2(L-(d-i))}\right),$$

with $\chi^2_{2(L-(d-i))}$ mutually independent chi-squared distributions, with $2(L-(d-i))$ degrees of freedom. Using that $\boldsymbol{E}[\log(\chi^2_\nu)] = \log(2) + \psi(\nu/2)$, it follows that:

$$\boldsymbol{E}[\log(\det(\Lambda))] \;=\; -d\log(L) + \sum_{i=1}^{d}\psi(L-(d-i)).$$

Following Smith (2000), $\boldsymbol{E}[\vec{q}_i\vec{q}_i^{\,*}] = d^{-1}\mathrm{Id}$, thus by eq.(8.10):

$$\boldsymbol{E}[\mathrm{Log}(Y)] \;=\; \left(-\log(L) + \frac{1}{d}\sum_{i=1}^{d}\psi(L-(d-i))\right)\cdot\mathrm{Id} \;=\; c(d,L)\cdot\mathrm{Id}.$$

Plugging this back into eq.(8.9) yields $b(X, f) = c(d, L) \cdot f$.

For the second part of the theorem, observe that $\widetilde{X}_\ell$ $(1 \le \ell \le n)$ is unbiased with respect to f, since:

$$\begin{aligned} b(\widetilde{X}_\ell, f) &= f^{1/2} * \boldsymbol{E}[\mathrm{Log}(f^{-1/2} * \widetilde{X}_\ell)] \\ &= f^{1/2} * \boldsymbol{E}[\mathrm{Log}(e^{-c(d,L)}\mathrm{Id}) + \mathrm{Log}(f^{-1/2} * X_\ell)] \\ &= f^{1/2} * (-c(d,L)\mathrm{Id} + c(d,L)\mathrm{Id}) \;=\; \mathbf{0}, \end{aligned}$$

using that $\mathrm{Log}(AB) = \mathrm{Log}(A) + \mathrm{Log}(B)$ for commuting matrices A, B, and $\boldsymbol{E}[\mathrm{Log}(f^{-1/2} * X_\ell)] = c(d, L) \cdot \mathrm{Id}$ as shown above. By eq.(7.13), the unique intrinsic mean of $\widetilde{X}_\ell$ on $\mathcal{M}$ is characterized by f such that $b(\widetilde{X}_\ell, f) = \boldsymbol{E}[\mathrm{Log}_f(\widetilde{X}_\ell)] = \mathbf{0}$, i.e., f is the unique intrinsic mean of $\widetilde{X}_\ell$ for each $\ell = 1, \ldots, n$. Since the distribution of $\widetilde{X}_\ell$ has finite second moment (rescaled complex Wishart distribution), the convergence in probability follows by Proposition 3.1. $\square$

8.6 Proofs of Proposition 4.2 and Lemma 4.3

Proof. In this proof, we directly derive the stronger general linear congruence equivariance property in Lemma 4.3. The weaker unitary congruence equivariance property in Proposition 4.2 then follows directly by substituting wavelet thresholding or shrinkage of coefficients that is only equivariant under unitary congruence transformation, (instead of trace thresholding as in Lemma 4.3, which is equivariant under general linear congruence transformation of the coefficients).

Let $M^X_{j,k}$, $M^{\hat{f}}_{j,k}$, $D^X_{j,k}$ and $D^{\hat{f}}_{j,k}$ be the midpoints and wavelet coefficients at scale-location (j, k) based on the observations $(X_\ell)_\ell$ and the estimator $(\hat{f}_\ell)_\ell$ respectively. Analogously, let $M^{X,A}_{j,k}$, $M^{\hat{f},A}_{j,k}$, $D^{X,A}_{j,k}$ and $D^{\hat{f},A}_{j,k}$ be the midpoints and wavelet coefficients based on the observations $(A * X_\ell)_\ell$ and the estimator $(A * \hat{f}_\ell)_\ell$ respectively, where here and throughout this proof $A \in \mathrm{GL}(d, \mathbb{C})$. Below, we repeatedly make use of the identities $A * \mathrm{Exp}_M(H) = \mathrm{Exp}_{A*M_1}(A*H)$ and $A * \mathrm{Log}_{M_1}(M_2) = \mathrm{Log}_{A*M_1}(A * M_2)$ for $M_1, M_2 \in \mathcal{M}$ and $H \in \mathcal{H}$. In particular, denoting $\mathrm{Mid}(M_1, M_2) := \eta(M_1, M_2, 1/2)$ for the geodesic midpoint, also,

$$\begin{aligned} A * \mathrm{Mid}(M_1, M_2) \;=\; A * \mathrm{Exp}_{M_1}\left(\frac{1}{2}\mathrm{Log}_{M_1}(M_2)\right) \;=\; \\ \mathrm{Exp}_{A*M_1}\left(\frac{1}{2}\mathrm{Log}_{A*M_1}(A * M_2)\right) \;=\; \mathrm{Mid}(A * M_1, A * M_2). \end{aligned}$$

By construction, the finest-scale midpoints satisfy $M^{X,A}_{J,k} = A * M^X_{J,k}$. Repeated application of the above identity then implies,

$$M^{X,A}_{j,k} \;=\; A * M^X_{j,k} \qquad \text{for all } j, k. \tag{8.11}$$

Furthermore, since the predicted midpoints $\widetilde{M}^{X,A}_{j,k}$ are weighted intrinsic means of $(M^{X,A}_{j-1,k'})_{k'}$ according to eq.(2.5) in the main document, the same relation holds for the predicted midpoints, i.e., $\widetilde{M}^{X,A}_{j,k} = A * \widetilde{M}^X_{j,k}$ for all j, k. Consequently, for the wavelet coefficients at each scale-location (j, k),

$$D^{X,A}_{j,k} \;=\; 2^{-j/2}\mathrm{Log}_{A*\widetilde{M}^X_{j,2k+1}}\big(A * M^X_{j,2k+1}\big) \;=\; A * D^X_{j,k}. \tag{8.12}$$

In Lemma 4.3, we threshold or shrink the wavelet coefficients based on the trace of the whitened coefficients, for which:

$$\begin{aligned}
\mathrm{Tr}(\mathfrak{D}_{j,k}^{X,A}) &= 2^{-j/2}\mathrm{Tr}\left(\mathrm{Log}\big((A*\widetilde{M}_{j,2k+1}^{X})^{-1/2}*(A*M_{j,2k+1}^{X})\big)\right) \\
&= 2^{-j/2}\left(\mathrm{Tr}\big(\mathrm{Log}(A*M_{j,2k+1}^{X})\big)-\mathrm{Tr}\big(\mathrm{Log}(A*\widetilde{M}_{j,2k+1}^{X})\big)\right) \\
&= 2^{-j/2}\left(\mathrm{Tr}\big(\mathrm{Log}(M_{j,2k+1}^{X})\big)-\mathrm{Tr}\big(\mathrm{Log}(\widetilde{M}_{j,2k+1}^{X})\big)\right) \\
&= \mathrm{Tr}(\mathfrak{D}_{j,k}^{X}), \qquad (8.13)
\end{aligned}$$

using that $\mathrm{Tr}(\mathrm{Log}(A*X)) = \mathrm{Tr}(\mathrm{Log}(X)) + \mathrm{Tr}(\mathrm{Log}(AA^*))$ and $\mathrm{Tr}(\mathrm{Log}(X^t)) = t\mathrm{Tr}(\mathrm{Log}(X))$ for $X \in \mathcal{M}$ and $t \in \mathbb{R}$, which follows from the fact that $\mathrm{Tr}(\mathrm{Log}(X)) = \log(\det(X))$ and the properties of the determinant and ordinary logarithm. Let $g(\mathrm{Tr}(\mathfrak{D}_{j,k}^{X})) \in \mathbb{R}$ be a thresholding or shrinkage rule depending on $\mathrm{Tr}(\mathfrak{D}_{j,k}^{X})$, such that $D_{j,k}^{\hat{f}} = g(\mathrm{Tr}(\mathfrak{D}_{j,k}^{X}))D_{j,k}^{X}$. Due to the invariance in eq.(8.13) combined with eq.(8.12), it immediately follows that:

$$D_{j,k}^{\hat{f},A} \;=\; g(\mathrm{Tr}(\mathfrak{D}_{j,k}^{X,A}))D_{j,k}^{X,A} \;=\; A*\big(g(\mathrm{Tr}(\mathfrak{D}_{j,k}^{X}))D_{j,k}^{X}\big) \;=\; A*D_{j,k}^{\hat{f}} \qquad \text{for all } j,k.$$

The wavelet-thresholded estimator $(\hat{f}_\ell)_\ell$ is retrieved via the inverse wavelet transform applied to the set of thresholded wavelet coefficients (and coarse-scale midpoints). At scale $j = 0$, by eq.(8.11), $M_{0,k}^{\hat{f},A} = M_{0,k}^{X,A} = A*M_{0,k}^{X} = A*M_{0,k}^{\hat{f}}$. At the odd locations $2k+1$ at the next coarser scale $j = 1$,

$$\begin{aligned}
M_{1,2k+1}^{\hat{f},A} &= \mathrm{Exp}_{\widetilde{M}_{1,2k+1}^{\hat{f},A}}\left(2^{1/2}D_{j,k}^{\hat{f},A}\right) \\
&= \mathrm{Exp}_{A*\widetilde{M}_{1,2k+1}^{\hat{f}}}\left(A*\left(2^{1/2}D_{j,k}^{\hat{f}}\right)\right) \\
&= A*\mathrm{Exp}_{\widetilde{M}_{1,2k+1}^{\hat{f}}}\left(2^{1/2}D_{j,k}^{\hat{f}}\right) \\
&= A*M_{1,2k+1}^{\hat{f}},
\end{aligned}$$

using that $\widetilde{M}_{1,2k+1}^{\hat{f},A} = A*\widetilde{M}_{1,2k+1}^{\hat{f}}$, since the same relation holds for $(M_{0,k'}^{\hat{f},A})_{k'}$ and the predicted midpoints are weighted intrinsic means of $(M_{0,k'}^{\hat{f},A})_{k'}$. Also, at the even locations $2k$,

$$\begin{aligned}
M_{1,2k}^{\hat{f},A} &= M_{0,k}^{\hat{f},A}*\big(M_{1,2k+1}^{\hat{f},A}\big)^{-1} \\
&= (A*M_{0,k}^{\hat{f}})*\big(A*M_{1,2k+1}^{\hat{f}}\big)^{-1} \\
&= A*\left(M_{0,k}^{\hat{f}}*\big(M_{1,2k+1}^{\hat{f}}\big)^{-1}\right) \\
&= A*M_{1,2k}^{\hat{f}}.
\end{aligned}$$

Iterating the same argument up to the finest scale $j = J$ yields the desired result $\hat{f}_{A,\ell} = A*\hat{f}_\ell$ for each $\ell = 1,\ldots,2^J$. □

8.7 Proof of Proposition 4.4

Proof. Let us write $M^X_{J,k-1} := X_k = f_k^{1/2} * W_k$ for $k = 1, \ldots, n$, where the distribution of W_k does not depend on f_k, and the intrinsic mean of W_k is the identity Id. The latter follows from the fact that X_k has intrinsic mean f_k, since:

$$\begin{aligned} \boldsymbol{E}[\mathrm{Log}_{\mathrm{Id}}(W_k)] &= \boldsymbol{E}[f_k^{-1/2} * \mathrm{Log}_{f_k}(f_k^{1/2} * W_k)] \\ &= f_k^{-1/2} * \boldsymbol{E}[\mathrm{Log}_{f_k}(X_k)] \\ &= f_k^{-1/2} * \mathbf{0} \;=\; \mathbf{0}, \end{aligned}$$

and the intrinsic mean μ of W_k is uniquely characterized by $\boldsymbol{E}[\mathrm{Log}_\mu(W_k)] = \mathbf{0}$. First, we verify that:

$$\mathrm{Tr}(\mathrm{Log}(M^X_{j,k})) \;=\; \mathrm{Tr}(\mathrm{Log}(M^f_{j,k})) + \mathrm{Tr}(\mathrm{Log}(M^W_{j,k})) \qquad \text{for all } j,k, \tag{8.14}$$

where $M^X_{j,k}$, $M^f_{j,k}$, and $M^W_{j,k}$ are the midpoints at scale-location (j,k) based on the sequences $(X_\ell)_\ell$, $(f_\ell)_\ell$, and $(W_\ell)_\ell$ respectively. For convenience, as before, denote $\mathrm{Mid}(X_1, X_2) := \eta(M_1, M_2, 1/2)$ for the geodesic midpoint. Using that $\mathrm{Tr}(\mathrm{Log}(AB)) = \mathrm{Tr}(\mathrm{Log}(A)) + \mathrm{Tr}(\mathrm{Log}(B))$ and $\mathrm{Log}(A^t) = t\mathrm{Log}(A)$ for any $A, B \in \mathcal{M}$, decompose:

$$\begin{aligned} \mathrm{Tr}(\mathrm{Log}(M^X_{j,k})) &= \mathrm{Tr}(\mathrm{Log}(\mathrm{Mid}(M^X_{j+1,2k}, M^X_{j+1,2k+1}))) \\ &= \mathrm{Tr}\big(\mathrm{Log}\big((M^X_{j+1,2k})^{1/2} * \big((M^X_{j+1,2k})^{-1/2} * M^X_{j+1,2k+1}\big)^{1/2}\big)\big) \\ &= \frac{1}{2}\mathrm{Tr}(\mathrm{Log}(M^X_{j+1,2k})) + \frac{1}{2}\mathrm{Tr}(\mathrm{Log}(M^X_{j+1,2k+1})) \\ &\;\;\vdots \\ &= \frac{1}{2^{J-j}} \sum_{\ell=0}^{2^{J-j}-1} \mathrm{Tr}(\mathrm{Log}(M^X_{J,(2k)^{J-j-1}+\ell})) \\ &= \frac{1}{2^{J-j}} \sum_{\ell=0}^{2^{J-j}-1} \mathrm{Tr}(\mathrm{Log}(f_{(2k)^{J-j-1}+\ell+1})) \\ &\quad + \frac{1}{2^{J-j}} \sum_{\ell=0}^{2^{J-j}-1} \mathrm{Tr}(\mathrm{Log}(W_{(2k)^{J-j-1}+\ell+1})) \\ &\;\;\vdots \\ &= \mathrm{Tr}(\mathrm{Log}(\mathrm{Mid}(M^f_{j+1,2k}, M^f_{j+1,2k+1}))) + \mathrm{Tr}(\mathrm{Log}(\mathrm{Mid}(M^W_{j+1,2k}, M^W_{j+1,2k+1}))) \\ &= \mathrm{Tr}(\mathrm{Log}(M^f_{j,k})) + \mathrm{Tr}(\mathrm{Log}(M^W_{j,k})). \end{aligned}$$

Second, we also verify that for each scale j and location k,

$$\mathrm{Tr}(\mathrm{Log}(\widetilde{M}^X_{j,2k+1})) \;=\; \mathrm{Tr}(\mathrm{Log}(\widetilde{M}^f_{j,2k+1})) + \mathrm{Tr}(\mathrm{Log}(\widetilde{M}^W_{j,2k+1})), \tag{8.15}$$

where $\widetilde{M}^X_{j,k'}$, $\widetilde{M}^f_{j,k'}$, and $\widetilde{M}^W_{j,k'}$ are the imputed midpoints at scale-location (j,k') based on the sequences $(X_\ell)_\ell$, $(f_\ell)_\ell$, and $(W_\ell)_\ell$ respectively. By eq.(2.5) in the main document, the predicted midpoints at the odd locations $2k+1$ satisfy:

$$\widetilde{M}^X_{j,2k+1} \;=\; \mathrm{Exp}_{\widetilde{M}^X_{j,2k+1}}\left(\sum_{\ell=-L}^{L} C_{N,2\ell+N}\mathrm{Log}_{\widetilde{M}^X_{j,2k+1}}(M^X_{j-1,k+\ell})\right),$$

with weights $\boldsymbol{C}_N = (C_{N,i})_{i=0,\ldots,2N-1}$ as in eq.(2.5). Here, without loss of generality we consider prediction away from the boundary, (at the boundary the sum runs over the $N = 2L+1$ closest available neighbors to $M_{j,k}$). Using eq.(8.14), we decompose,

$$\begin{aligned}
\mathrm{Tr}(\mathrm{Log}(\widetilde{M}^X_{j,2k+1})) &= \mathrm{Tr}\Big(\mathrm{Log}\Big(\mathrm{Exp}_{\widetilde{M}^X_{j,2k+1}}\Big(\sum_\ell C_{N,2\ell+N}\mathrm{Log}_{\widetilde{M}^X_{j,2k+1}}(M^X_{j-1,k+\ell})\Big)\Big)\Big)\\
&= \mathrm{Tr}(\mathrm{Log}(\widetilde{M}^X_{j,2k+1})) + \mathrm{Tr}\left((\widetilde{M}^X_{j,2k+1})^{-1/2} * \Big(\sum_\ell C_{N,2\ell+N}\mathrm{Log}_{\widetilde{M}^X_{j,2k+1}}(M^X_{j-1,k+\ell})\Big)\right)\\
&= \mathrm{Tr}(\mathrm{Log}(\widetilde{M}^X_{j,2k+1})) + \mathrm{Tr}\left(\sum_\ell C_{N,2\ell+N}\mathrm{Log}\Big((\widetilde{M}^X_{j,2k+1})^{-1/2} * M^X_{j-1,k+\ell}\Big)\right)\\
&= \mathrm{Tr}(\mathrm{Log}(\widetilde{M}^X_{j,2k+1})) + \sum_\ell C_{N,2\ell+N}\Big(\mathrm{Tr}(\mathrm{Log}(M^X_{j-1,k+\ell})) - \mathrm{Tr}(\mathrm{Log}(\widetilde{M}^X_{j,2k+1}))\Big)\\
&= \sum_\ell C_{N,2\ell+N}\mathrm{Tr}(\mathrm{Log}(M^X_{j-1,k+\ell}))\\
&= \sum_\ell C_{N,2\ell+N}\mathrm{Tr}(\mathrm{Log}(M^f_{j-1,k+\ell})) + \sum_\ell C_{N,2\ell+N}\mathrm{Tr}(\mathrm{Log}(M^W_{j-1,k+\ell}))\\
&\vdots\\
&= \mathrm{Tr}(\mathrm{Log}(\widetilde{M}^f_{j,2k+1})) + \mathrm{Tr}(\mathrm{Log}(\widetilde{M}^W_{j,2k+1})),
\end{aligned}$$

where we used in particular $g*\mathrm{Log}_{X_1}(X_2) = \mathrm{Log}_{g*X_1}(g*X_2)$ and $g*\mathrm{Exp}_{X_1}(X_2) = \mathrm{Exp}_{g*X_1}(g*X_2)$ for any $g \in \mathrm{GL}(d,\mathbb{C})$, and the fact that $\sum_\ell C_{N,2\ell+N} = 1$.

The first claim in the Proposition now follows from eq.(8.14) and eq.(8.15) through:

$$\begin{aligned}
\mathrm{Tr}(\mathfrak{D}^X_{j,k}) &= 2^{-j/2}\mathrm{Tr}\Big(\mathrm{Log}\Big((\widetilde{M}^X_{j,2k+1})^{-1/2} * M^X_{j,2k+1}\Big)\Big)\\
&= 2^{-j/2}\Big(\mathrm{Tr}(\mathrm{Log}(M^X_{j,2k+1})) - \mathrm{Tr}(\mathrm{Log}(\widetilde{M}^X_{j,2k+1}))\Big)\\
&= 2^{-j/2}\mathrm{Tr}(\mathrm{Log}(M^f_{j,2k+1})) + 2^{-j/2}\mathrm{Tr}(\mathrm{Log}(M^W_{j,2k+1}))\\
&\quad -2^{-j/2}\Big(\mathrm{Tr}(\mathrm{Log}(\widetilde{M}^f_{j,2k+1})) + \mathrm{Tr}(\mathrm{Log}(\widetilde{M}^W_{j,2k+1}))\Big)\\
&= \mathrm{Tr}(\mathfrak{D}^f_{j,k}) + \mathrm{Tr}(\mathfrak{D}^W_{j,k}).
\end{aligned} \tag{8.16}$$

For the second claim in the Proposition, first observe:

$$\boldsymbol{E}[\mathrm{Tr}(\mathrm{Log}(M^W_{j,k}))] \;=\; \frac{1}{2^{J-j}}\sum_{\ell=0}^{2^{J-j}-1}\boldsymbol{E}[\mathrm{Tr}(\mathrm{Log}(W_{(2k)^{J-j-1}+\ell+1}))] \;=\; 0, \qquad \text{for each } j,k,$$

using that $\boldsymbol{E}[\mathrm{Tr}(\mathrm{Log}(W_\ell))] = 0$ for each $\ell = 1,\ldots,n$, which is implied by $\boldsymbol{E}[\mathrm{Log}_{\mathrm{Id}}(W_\ell)] = \mathbf{0}$. As a consequence, also,

$$\boldsymbol{E}[\mathrm{Tr}(\mathrm{Log}(\widetilde{M}^W_{j,2k+1}))] \;=\; \sum_\ell C_{N,2\ell+N}\mathrm{Tr}(\mathrm{Log}(M^W_{j-1,k+\ell})) \;=\; 0, \qquad \text{for each } j,k,$$

and therefore,

$$\begin{aligned}
\boldsymbol{E}[\mathrm{Tr}(\mathfrak{D}^X_{j,k})] &= \mathrm{Tr}(\mathfrak{D}^f_{j,k}) + \boldsymbol{E}[\mathrm{Tr}(\mathfrak{D}^W_{j,k})]\\
&= \mathrm{Tr}(\mathfrak{D}^f_{j,k}) + 2^{-j/2}\boldsymbol{E}\Big[\mathrm{Tr}(\mathrm{Log}(M^W_{j,2k+1})) - \mathrm{Tr}(\mathrm{Log}(\widetilde{M}^W_{j,2k+1}))\Big]\\
&= \mathrm{Tr}(\mathfrak{D}^f_{j,k}).
\end{aligned}$$

For the variance of $\mathrm{Tr}(\mathfrak{D}^X_{j,k})$, we first note that the random variables $(W_\ell)_{\ell=1,\dots,n}$ are i.i.d., implying that the random variables $(\mathrm{Tr}(\mathrm{Log}(M^W_{j,k}))_{k=0,\dots,2^j-1}$ on scale j are independent with equal variance. We write out:

$$\begin{aligned}
\mathrm{Var}(\mathrm{Tr}(\mathfrak{D}^X_{j,k})) &= 2^{-j}\mathrm{Var}\left(\mathrm{Tr}(\mathrm{Log}(M^W_{j,2k+1})) - \mathrm{Tr}(\mathrm{Log}(\widetilde{M}^W_{j,2k+1}))\right) \\
&= 2^{-j}\mathrm{Var}\Big(\mathrm{Tr}(\mathrm{Log}(M^W_{j,2k+1})) - \sum_\ell C_{L,2\ell+N}\mathrm{Tr}(\mathrm{Log}(M^W_{j-1,k+\ell}))\Big) \\
&= 2^{-j}\mathrm{Var}\Big(\mathrm{Tr}(\mathrm{Log}(M^W_{j,2k+1})) - C_{N,N}\mathrm{Tr}(\mathrm{Log}(M^W_{j-1,k}))\Big) \\
&\quad + \; 2^{-j}\sum_{-L\le\ell\le L;\ell\neq 0} C^2_{N,2\ell+N}\mathrm{Var}(\mathrm{Tr}(\mathrm{Log}(M^W_{j-1,k+\ell}))) \\
&= 2^{-(j+1)}\mathrm{Var}(\mathrm{Tr}(\mathrm{Log}(M^W_{j,2k}))) \\
&\quad + \; 2^{-j}\Big(\sum_\ell C^2_{N,2\ell+N} - 1\Big)\mathrm{Var}(\mathrm{Tr}(\mathrm{Log}(M^W_{j-1,k+\ell}))) \\
&= 2^{-(j+1)}\sum_\ell C^2_{N,2\ell+N}\mathrm{Var}(\mathrm{Tr}(\mathrm{Log}(M^W_{j,0}))), \qquad (8.17)
\end{aligned}$$

where in the final two steps we used that $C_{N,N} = 1$, and by the independence of the midpoints within each scale, for each k,

$$\begin{aligned}
\mathrm{Var}(\mathrm{Tr}(\mathrm{Log}(M^W_{j-1,k}))) &= \mathrm{Var}\left(\frac{1}{2}\mathrm{Tr}(\mathrm{Log}(M^W_{j,2k})) + \frac{1}{2}\mathrm{Tr}(\mathrm{Log}(M^W_{j,2k+1}))\right) \\
&= \frac{1}{2}\mathrm{Var}(\mathrm{Tr}(\mathrm{Log}(M^W_{j,0}))).
\end{aligned}$$

It remains to derive an expression for $\mathrm{Var}(\mathrm{Tr}(\mathrm{Log}(M^W_{j,0})))$. By repeated application of the above argument,

$$\begin{aligned}
\mathrm{Var}(\mathrm{Tr}(\mathrm{Log}(M^W_{j,0}))) &= \frac{1}{2^{J-j}}\mathrm{Var}(\mathrm{Tr}(\mathrm{Log}(M^W_{J,0}))) \\
&= \frac{1}{2^{J-j}}\mathrm{Var}(\mathrm{Tr}(\mathrm{Log}(W_1))), \qquad (8.18)
\end{aligned}$$

with $W_1 \sim W^c_d(L, L^{-1}e^{-c(d,L)}\mathrm{Id})$. As in the proof of Theorem 4.1,

$$\mathrm{Tr}(\mathrm{Log}(W_1)) \sim -d\log(2e^{c(d,L)}L) + \sum_{i=1}^d \log\left(\chi^2_{2(L-(d-i))}\right).$$

The variance of a $\log(\chi^2_\nu)$ distribution equals $\psi'(\nu/2)$, (with $\psi'(\cdot)$ the trigamma function), therefore:

$$\mathrm{Var}(\mathrm{Tr}(\mathrm{Log}(W_1))) = \sum_{i=1}^d \psi'(L-(d-i)).$$

Combining the above result with eq.(8.17) and eq.(8.18) finishes the proof. □

8.8 Proof of Corollary 4.5

Proof. Analogous to the proof of Theorem 4.1, $W_1, \ldots, W_n \overset{\text{iid}}{\sim} W_d^c(L, L^{-1}e^{-c(d,L)}\text{Id})$ are unitarily invariant, see (Muirhead, 1982, Section 3.2). By the same argument as in eq.(8.11) the repeated midpoints based on unitarily invariant random variables satisfy $U * M_{j,k}^W \overset{d}{=} M_{j,k}^W$ for each j, k and $U \in \mathcal{U}_d$. It follows that the predicted midpoints $\widetilde{M}_{j,2k+1}^W$ are unitarily invariant as well, as they can be expressed as weighted intrinsic averages of the midpoints $(M_{j-1,k}^W)_k$, which are unitarily invariant themselves. That is, $U * \widetilde{M}_{j,2k+1}^W \overset{d}{=} \widetilde{M}_{j,2k+1}^W$ for each j, k and $U \in \mathcal{U}_d$. Combining the above results, it follows that the random whitened coefficient $\mathfrak{D}_{j,k}^W$ is unitarily invariant, as for each $U \in \mathcal{U}_d$,

$$\begin{aligned} U * \mathfrak{D}_{j,k}^W &= U * \text{Log}\left((\widetilde{M}_{j,2k+1}^W)^{-1/2} * M_{j,2k+1}\right) \\ &= \text{Log}\left((U * \widetilde{M}_{j,2k+1}^W)^{-1/2} * (U * M_{j,2k+1})\right) \\ &\overset{d}{=} \text{Log}\left((\widetilde{M}_{j,2k+1}^W)^{-1/2} * M_{j,2k+1}\right) \\ &= \mathfrak{D}_{j,k}^W, \end{aligned}$$

using that $U * \text{Log}(X) = \text{Log}(U * X)$ for $U \in \mathcal{U}_d$. By the same argument as in Theorem 4.1, if we write the eigendecomposition $\mathfrak{D}_{j,k}^W = Q * \Lambda$, then for a unitarily invariant random matrix $\mathfrak{D}_{j,k}^W$,

$$\begin{aligned} \boldsymbol{E}[\mathfrak{D}_{j,k}^W] &= \boldsymbol{E}\left[\sum_{i=1}^{d} \lambda_i \vec{q}_i \vec{q}_i^*\right] \\ &= \boldsymbol{E}[\vec{q}_i \vec{q}_i^*]\boldsymbol{E}[\text{Tr}(\Lambda)] \\ &= \boldsymbol{E}[\vec{q}_i \vec{q}_i^*]\boldsymbol{E}[\text{Tr}(\mathfrak{D}_{j,k}^W)] \;=\; \mathbf{0}. \end{aligned}$$

Here we used that $\text{Tr}(Q * \Lambda) = \text{Tr}(\Lambda)$, since Q is a unitary matrix ($\mathfrak{D}_{j,k}^W$ is Hermitian), combined with the result $\boldsymbol{E}[\text{Tr}(\mathfrak{D}_{j,k}^W)] = 0$ in Proposition 4.4. □

9 Appendix III: Additional details Section 5.1

Estimation procedures Section 5.1 This appendix section provides more details on the matrix curve estimation procedures considered in the simulated data scenarios in Section 5.1 in the main document. Each estimation procedure takes as input an initial dyadic sequence of random HPD matrix-valued observations $X_1, \ldots, X_n \in \mathcal{M}$ observed on an equidistant grid $t_1, \ldots, t_n \in \mathbb{R}$ and outputs a denoised sequence of HPD matrix-valued observations $\hat{f}(t_1), \ldots, \hat{f}(t_n) \in \mathcal{M}$.

- **Linear wavelet thresholding**: the input data $X_1, \ldots, X_n$ is transformed to the intrinsic wavelet domain by means of the forward average-interpolating wavelet transform of Section 2 in the main document subject to respectively the Riemannian, Log-Euclidean or Cholesky metric, and all wavelet coefficients at scales $j > J_0$ are set to zero. The smoothed curve estimate $\hat{f}(t_1), \ldots, \hat{f}(t_n)$ is obtained by application of the intrinsic backward average-interpolating wavelet transform. The main tuning parameter in the case of

Table 1: Estimation procedure metrics and their properties.

Metric	U-equiv.*	A-equiv.†	PD Estimates	Wishart B-C**
Riemannian	✓	✓	✓	✓
Log-Euclidean	✓	✗	✓	✗
Cholesky	✗	✗	✓	✓
Euclidean	✓	✗	✗	✓

$*, \dagger$: U-equiv. and A-equiv. respectively denote whether the estimator is equivariant under congruence transformation by a unitary matrix $U \in \mathcal{U}_d$ or a general linear matrix $A \in \mathrm{GL}(\mathbb{C}, d)$, see Section 4.1.

$**$: Wishart B-C denotes whether a bias-correction (B-C) is available in the context of spectral matrix estimation, where the periodogram data is asymptotically Wishart distributed.

linear wavelet thresholding is the maximum scale of nonzero wavelet coefficients J_0. The impact of the average-interpolation order of the wavelet transform is small in terms of the estimation error compared to the choice of the scale parameter J_0. For this reason the refinement order is fixed at $N = 5$ for all simulated scenarios in Section 5. Linear wavelet thresholding is implemented in the `pdSpecEst`-package by the function `pdSpecEst1D()` with arguments `alpha = 0`, `jmax` set to the maximum scale of nonzero coefficients J_0, and `metric` set to metric considered for estimation.

- **Nonlinear wavelet thresholding**: the input data $X_1, \ldots, X_n$ is transformed to the intrinsic wavelet domain the same way as for the linear wavelet thresholding procedure. The nonlinear wavelet thresholding procedure considers dyadic tree-structured thresholding based on the traces of the individual coefficients by minimizing the complexity penalized loss criterion given in eq.(5.1) and explained in more detail in the main document. The main tuning parameter is the regularization parameter $\lambda \geq 0$, and the refinement order of the wavelet transforms is fixed at $N = 5$ for all simulation scenarios equivalent to the linear thresholding procedure. For sufficiently large n, the scalar coefficients $d_{j,k}$ are approximately normally distributed at reasonably coarse scales j, as the scalar coefficients $d_{j,k}$ are essentially locally weighted averages of the observations. For normally distributed coefficients, a natural choice for the regularization parameter is the universal threshold $\lambda \sim \sigma_w \sqrt{2 \log(n)}$, with n the total number of wavelet coefficients and σ_w^2 the noise variance determined either via eq.(4.1) in the main document or from the data itself. Tree-structured trace thresholding is implemented in the `pdSpecEst`-package by the function `pdSpecEst1D()` with arguments `alpha = 1` to use a universal threshold multiplied by $\alpha = 1$, and `metric` set to the metric considered for estimation.

- **Nearest-Neighbor (NN) regression**: intrinsic nearest-neighbor regression is implemented by replacing ordinary local Euclidean averages by their intrinsic counterparts based on the Riemannian, Log-Euclidean and Cholesky metric using the function `pdMean()` in the `pdSpecEst`-package. In the case of the Riemannian metric, the local intrinsic averages are calculated efficiently by the gradient descent algorithm in Pennec (2006). The main tuning parameter in this benchmark procedure is the number of nearest neighbors used in the local averages.

- **Cubic Spline (CS) regression**: intrinsic cubic smoothing spline regression is implemented in the space of HPD matrices based on the Riemannian, Log-Euclidean and

Cholesky metric. For the Riemannian metric, we implemented the penalized regression approach in Boumal and Absil (2011a) and Boumal and Absil (2011b), with penalty parameters $(\lambda = 0, \mu > 0)$, such that the minimizers of the objective function are approximate cubic splines in the manifold of HPD matrices. The Riemannian conjugate gradient descent method in Boumal and Absil (2011b) to compute the estimator is available through the function `pdSplineReg()` in the `pdSpecEst`-package. Here, we use a backtracking line search based on the Armijo-Goldstein condition. The main tuning parameter in this benchmark procedure is the regularization parameter in the penalized loss criterion.

- **Local polynomial (LP) regression**: intrinsic local polynomial regression of degree $p = 0$ (LP-0) and degree $p = 3$ (LP-3) respectively is implemented in the space of HPD matrices based on the Riemannian metric, Log-Euclidean metric and Cholesky metric. For the Riemannian metric, we have only implemented the locally constant estimator, i.e. degree $p = 0$, as local polynomial regression under the Riemannian metric for $p > 0$ requires the optimization of a non-convex objective function and is computationally quite challenging. We refer to Yuan et al. (2012) for additional details. The main tuning parameter in this benchmark procedure is the bandwidth parameter of the local polynomials.

- **Multitaper spectral estimation**: the multitaper benchmark estimator is only considered in the periodogram noise scenario given in Table 2, as this is the only simulated scenario that provides input time series data in addition to the input (periodogram) observations $X_1, \ldots, X_n$. The multitaper spectral estimate takes as input the generated d-dimensional stationary time trace and is based on $L \geq d$ discrete prolate spheroidal (DPSS) taper functions using the function `pdPgram()`, thereby guaranteeing an HPD matrix curve estimate $\hat{f}(t_1), \ldots, \hat{f}(t_n) \in \mathcal{M}$. The main tuning parameter in this benchmark procedure is the number of DPSS tapers L.

References

Bhatia, R. (2009). *Positive Definite Matrices*. New Jersey: Princeton University Press.

Boumal, N. and P.-A. Absil (2011a). A discrete regression method on manifolds and its application to data on SO(n). *IFAC Proceedings Volumes 44*(1), 2284–2289.

Boumal, N. and P.-A. Absil (2011b). Discrete regression methods on the cone of positive-definite matrices. In *IEEE ICASSP, 2011*, pp. 4232–4235.

do Carmo, M. (1992). *Riemannian Geometry*. Boston: Birkhäuser.

Higham, N. J. (2008). *Functions of Matrices: Theory and Computation*. Philadelphia: Siam.

Ho, J., G. Cheng, H. Salehian, and B. Vemuri (2013). Recursive Karcher expectation estimators and recursive law of large numbers. In *AISTATS, 2013*, pp. 325–332.

Jeuris, B., R. Vandebril, and B. Vandereycken (2012). A survey and comparison of contemporary algorithms for computing the matrix geometric mean. *Electronic Transactions on Numerical Analysis 39*, 379–402.

Le, H. (1995). Mean size-and-shapes and mean shapes: a geometric point of view. *Advances in Applied Probability 27*(1), 44–55.

Muirhead, R. (1982). *Aspects of Multivariate Statistical Theory*. New Jersey: John Wiley & Sons.

Pennec, X. (2006). Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision 25*(1), 127–154.

Pennec, X., P. Fillard, and N. Ayache (2006). A Riemannian framework for tensor computing. *International Journal of Computer Vision 66*(1), 41–66.

Skovgaard, L. (1984). A Riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics 11*(4), 211–223.

Smith, S. (2000). Intrinsic Cramér-Rao bounds and subspace estimation accuracy. In *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop*, pp. 489–493. IEEE.

Tulino, A. and S. Verdú (2004). *Random Matrix Theory and Wireless Communications*. Hanover: Now Publishers Inc.

Villani, C. (2009). *Optimal Transport: Old and New*. Berlin: Springer-Verlag.

Yuan, Y., H. Zhu, W. Lin, and J. Marron (2012). Local polynomial regression for symmetric positive definite matrices. *Journal of the Royal Statistical Society: Series B 74*(4), 697–719.