INTRINSIC WAVELET REGRESSION FOR CURVES OF HERMITIAN POSITIVE DEFINITE MATRICES

Chau, J. and Rainer von Sach

REPRINT | 2019 / 51

UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

Intrinsic wavelet regression for curves of Hermitian positive definite matrices

Joris Chau* and Rainer von Sachs

Abstract

Intrinsic wavelet transforms and wavelet estimation methods are introduced for curves in the non-Euclidean space of Hermitian positive definite matrices, with in mind the application to Fourier spectral estimation of multivariate stationary time series. The main focus is on intrinsic average-interpolation wavelet transforms in the space of positive definite matrices equipped with an affine-invariant Riemannian metric, and convergence rates of linear wavelet thresholding are derived for intrinsically smooth curves of Hermitian positive definite matrices. In the context of multivariate Fourier spectral estimation, intrinsic wavelet thresholding is equivariant under a change of basis of the time series, and nonlinear wavelet thresholding is able to capture localized features in the spectral density matrix across frequency, always guaranteeing positive definite estimates. The finite-sample performance of intrinsic wavelet thresholding is assessed by means of simulated data and compared to several benchmark estimators in the Riemannian manifold. Further illustrations are provided by examining the multivariate spectra of trial-replicated brain signal time series recorded during a learning experiment.

Keywords: Riemannian manifold, Hermitian positive definite matrices, Intrinsic wavelet transform, Wavelet thresholding, Fourier spectral matrix, Multivariate time series.

1 Introduction

In multivariate time series analysis, the second-order behavior of a multivariate time series is studied by means of its autocovariance matrices in the time domain, or its spectral density matrices in the frequency domain. Non-degenerate spectral density matrices are necessarily curves of Hermitian positive definite (HPD) matrices, and one generally constrains a spectral curve estimator to preserve these properties. This is important for several reasons: i) interpretation of the spectral estimator as the Fourier transform of symmetric positive definite (SPD) autocovariance matrices in the time domain or as HPD covariance matrices across frequency in the Fourier domain; ii) well-defined transfer functions in the Cramér representation of the

*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.

time series for the purpose of e.g. simulation of time series and bootstrapping; iii) sufficient regularity to avoid computational problems in subsequent inference procedures (requiring e.g., the inverse of the estimated spectrum). Our main contribution is the development of intrinsic wavelet transforms and nonparametric wavelet regression for curves in the non-Euclidean space of HPD matrices, exploiting the geometric structure of the space as a Riemannian manifold. The primary focus is on nonparametric spectral density matrix estimation of stationary multivariate time series, but we emphasize that the methodology applies equally to general matrix-valued curve estimation or denoising problems, where the target is a curve of symmetric or Hermitian PD matrices. Examples include; curve denoising of SPD diffusion covariance matrices in diffusion tensor imaging as in e.g., Yuan et al. (2012), or estimation of time-varying autocovariance matrices of a locally stationary time series as in e.g., Dahlhaus (2012).

A first important consideration to perform estimation in the space of HPD matrices is the associated metric in the space. The metric gives the space its curvature and induces a distance between HPD matrices. Standard nonparametric spectral matrix estimation commonly relies on smoothing the periodogram via e.g., kernel regression as in (Brillinger, 1981, Chapter 5), (Brockwell and Davis, 2006, Chapter 11), or multitaper spectral estimation as in e.g, Walden (2000). These approaches equip the space of HPD matrices with the Euclidean (i.e., Frobenius) metric and view it as a *flat* space. An important disadvantage is that this metric space is incomplete, as the boundary of singular matrices lies at a finite distance. For this reason, flexible nonparametric (e.g., wavelet- or spline-) periodogram smoothing embedded in a Euclidean space cannot guarantee a PD spectral estimate. Exceptions to this rule include inflexible kernel or multitaper periodogram smoothing, which rely on a sufficiently large equivalent smoothing bandwidth for each matrix component. To avoid this issue, Dai and Guo (2004), Rosen and Stoffer (2007) and Krafty and Collinge (2013) among others construct an HPD spectral estimate as the square of an estimated curve of Cholesky square root matrices. This allows for more flexible estimation of the spectrum, such as individual smoothing of Cholesky matrix components, while at the same time guaranteeing an HPD spectral estimate. In this context, the space of HPD matrices is equipped with the Cholesky metric, where the distance between two matrices is given by the Euclidean distance between their Cholesky square roots. Unfortunately, the Cholesky metric and Cholesky-based smoothing are not equivariant to permutations of the components of the input time series. That is, if one reorders the time series components, the resulting spectral estimate is not necessarily a

permuted version of the spectral estimate obtained under the original input time series.

In this work, we exploit the geometric structure of the space of HPD matrices as a Riemannian manifold equipped with the affine-invariant (Pennec et al. (2006)) –also natural invariant (Smith (2000)), canonical (Holbrook et al. (2018)), trace (Yuan et al. (2012)), Rao-Fisher (Said et al. (2017))– Riemannian metric, or simply the Riemannian metric (Bhatia, 2009, Chapter 6), Dryden et al. (2009)). The affine-invariant Riemannian metric plays an important role in estimation problems in the space of symmetric or Hermitian PD matrices for several reasons: (i) the space of HPD matrices equipped with the Riemannian metric is a complete metric space, (ii) there is no swelling effect as with the Euclidean metric, where interpolating two HPD matrices may yield a matrix with a determinant larger than either of the original matrices (e.g., Pasternak et al. (2010)), and (iii) the induced Riemannian distance is invariant to congruence transformation by any invertible matrix, see Section 2. The first property guarantees an HPD spectral estimate, while allowing for flexible spectral matrix estimation as with Cholesky-based smoothing. The third property ensures that the spectral estimator is –not only– permutation or unitary congruence equivariant, but also *general linear congruence equivariant*, which essentially implies that the estimator does not nontrivially depend on the chosen coordinate system of the time series. In Dryden et al. (2009), the authors list several additional suitable metrics to perform estimation in the space of HPD matrices, one of which is the Log-Euclidean metric, also discussed in e.g., Yuan et al. (2012) or Boumal and Absil (2011b). The Log-Euclidean metric transforms the space of HPD matrices in a complete metric space and is unitary congruence invariant, but not general linear congruence invariant.

Several recent works on nonparametric curve regression in the space of SPD matrices equipped with the affine-invariant Riemannian metric include: intrinsic geodesic and linear regression in Pennec et al. (2006) and Zhu et al. (2009) among others, intrinsic local polynomial regression in Yuan et al. (2012) and intrinsic penalized spline-like regression in Boumal and Absil (2011b). In the context of frequency-specific spectral matrix estimation Holbrook et al. (2018) recently introduced a Bayesian geodesic Lagrangian Monte Carlo (gLMC) approach based on the affine-invariant Riemannian metric. The latter may not be best-suited to estimation of the entire spectral curve, as this requires application of the gLMC to each individual Fourier frequency, which is computationally quite challenging. In this work, we develop fast intrinsic wavelet regression in the manifold of HPD matrices equipped with the Riemannian metric.

Wavelet-based estimation of spectral matrices allows us to capture potentially very localized features, such as local peaks or troughs in the spectral matrix at pointwise frequencies or frequency bands, in contrast to the approaches mentioned above, which rely on globally homogeneous smoothness behavior. This paper is accompanied by an R-package `pdSpecEst` (**p**ositive **d**efinite **Spec**tral **Est**imation), which contains implementations of the presented material and is available on CRAN (Chau (2017)). The technical proofs and additional descriptions of the geometric notions and tools used in this paper can be found in the supplemental material.

2 Intrinsic AI wavelet transforms

We consider intrinsic wavelet transforms in the space of HPD matrices as generalizations of the average-interpolation (AI) wavelet transforms on the real line in Donoho (1993). In this sense, they are related to the *midpoint-interpolation* (MI) wavelet transforms in Rahman et al. (2005) for general symmetric Riemannian manifolds with tractable exponential and logarithmic maps. The MI approach in Rahman et al. (2005) projects manifold-valued input data to a set of tangent spaces and applies a Euclidean refinement scheme to the projected data. Such an approach might introduce a certain degree of ambiguity as the base points of the projecting tangent spaces are specified by the user and different base points may lead to different wavelet coefficients. In contrast, the intrinsic AI transforms implement a refinement scheme –intrinsic to the considered geometry– directly to the manifold-valued data itself, without first projecting the data to a set of Euclidean spaces. The primary advantage of such an intrinsic approach is that in contrast to the MI approach in Rahman et al. (2005), the AI refinement scheme of order $k \geq 0$ reproduces *intrinsic polynomial* curves up to order k as defined in Hinkle et al. (2014), whereas the MI refinement scheme reproduces only geodesic curves, i.e., intrinsic polynomials of order $k = 1$. This polynomial reproduction property is a necessary condition to derive wavelet coefficient decay and nonparametric estimation rates for (intrinsically) smooth curves of HPD matrices in Section 3, which are not readily available in the same context for the MI wavelet transforms in Rahman et al. (2005).

Preliminaries and notations The space of $(d \times d)$-dimensional Hermitian positive definite matrices $\mathbb{P}_{d\times d}$ is not a vector space due to its positive definite constraints, but it is an open subset of the vector space of Hermitian matrices $\mathbb{H}_{d\times d}$ and as such is also a smooth manifold, see e.g., do Carmo (1992). For every $p \in \mathbb{P}_{d\times d}$, the tangent space $T_p(\mathbb{P}_{d\times d})$ is identified by

Manifold:	$\mathbb{P}_{d\times d} := \{p \in \mathbb{C}^{d\times d} \,:\, p = p^* \text{ and } \vec{z}^* p \vec{z} > 0, \text{ for } \vec{z} \in \mathbb{C}^d, \vec{z} \neq \vec{0}\}$	
Tangent spaces:	$T_p(\mathbb{P}_{d\times d}) \cong \mathbb{H}_{d\times d} := \{h \in \mathbb{C}^{d\times d} \,:\, h = h^*\}$	
Riemannian metric:	$\langle h_1, h_2\rangle_p = \mathrm{Tr}((p^{-1/2} * h_1)(p^{-1/2} * h_2)),$	$h_1, h_2 \in T_p(\mathbb{P}_{d\times d})$
Distance:	$\delta_R(p_1, p_2) = \|\mathrm{Log}(p_1^{-1/2} * p_2)\|_F,$	$p_1, p_2 \in \mathbb{P}_{d\times d}$
Geodesics:	$\eta(p_1, p_2, t) = p_1^{1/2} * (p_1^{-1/2} * p_2)^t,$	$p_1, p_2 \in \mathbb{P}_{d\times d}, 0 \leq t \leq 1$
Exponential map:	$\mathrm{Exp}_p(h) = p^{1/2} * \mathrm{Exp}(p^{-1/2} * h),$	$p \in \mathbb{P}_{d\times d}, h \in T_p(\mathbb{P}_{d\times d})$
Logarithmic map:	$\mathrm{Log}_p(q) = p^{1/2} * \mathrm{Log}(p^{-1/2} * q),$	$p, q \in \mathbb{P}_{d\times d}$
Parallel transport:	$\Gamma_p^q(h) = p^{1/2} * (p^{-1/2} * q)^{1/2} * p^{-1/2} * h,$	$p, q \in \mathbb{P}_{d\times d}, h \in T_p(\mathbb{P}_{d\times d})$

Table 1: Geometric tools for the Riemannian manifold of $(d \times d)$-dimensional HPD matrices $(\mathbb{P}_{d\times d}, g_R)$, equipped with the affine-invariant Riemannian metric.

$\mathbb{H}_{d\times d}$, and as detailed 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 $\mathbb{P}_{d\times d}$. By (Bhatia, 2009, Theorem 6.1.6 and Prop. 6.2.2), the Riemannian manifold $(\mathbb{P}_{d\times d}, g_R)$, equipped with the affine-invariant metric, is *geodesically complete*, and the geodesic segment joining any two points $p_1, p_2 \in \mathbb{P}_{d\times d}$ is uniquely existing. Further, for each $p \in \mathbb{P}_{d\times d}$ the exponential map Exp_p and logarithmic map Log_p are global diffeomorphisms with as domains $T_p(\mathbb{P}_{d\times d})$ and $\mathbb{P}_{d\times d}$ respectively. The parameterizations of the geometric notions in the Riemannian manifold $(\mathbb{P}_{d\times d}, g_R)$ used throughout this work are summarized in Table 1. For more detailed descriptions, we refer to Appendix I in the supplementary material or (Chau, 2018, Chapter 2). Here and throughout this paper, $y^{1/2}$ always denotes the Hermitian square root matrix of $y \in \mathbb{P}_{d\times d}$, and we write $y * x$ for the matrix congruence transformation $y^* x y$, where y^* denotes the conjugate transpose of y. $\|\cdot\|_F$ is the matrix Frobenius norm, and $\mathrm{Exp}(\cdot)$ and $\mathrm{Log}(\cdot)$ denote the matrix exponential and the (principal) matrix logarithm. For convenience, the affine-invariant Riemannian metric is usually referred to simply as the Riemannian metric throughout this paper.

A random variable $X : \Omega \to \mathbb{P}_{d\times d}$ is a measurable function from a probability space $(\Omega, \mathcal{A}, \nu)$ to the measurable space $(\mathbb{P}_{d\times d}, \mathcal{B}(\mathbb{P}_{d\times d}))$, with $\mathcal{B}(\mathbb{P}_{d\times d})$ the Borel algebra in the complete separable metric space $(\mathbb{P}_{d\times d}, \delta_R)$. By $P(\mathbb{P}_{d\times d})$, we denote the set of all probability measures on $(\mathbb{P}_{d\times d}, \mathcal{B}(\mathbb{P}_{d\times d}))$ and $P_m(\mathbb{P}_{d\times d})$ denotes the subset of probability measures in $P(\mathbb{P}_{d\times d})$ that have finite moments of order m with respect to the Riemannian distance, i.e., the L^m-Wasserstein space, see (Villani, 2009, Definition 6.4). In the intrinsic AI refinement scheme described below, the center of a random variable $X \sim \nu$ is characterized by its intrinsic (also Karcher or

Fréchet) mean. The set of intrinsic means is given by the points that minimize the second moment with respect to the Riemannian distance δ_R,

$$\mu \;=\; \mathbb{E}_\nu[X] \;:=\; \arg\min_{y\in\mathrm{supp}(\nu)} \int_{\mathbb{P}_{d\times d}} \delta_R(y,x)^2 \;\nu(dx).$$

If $\nu \in P_2(\mathbb{P}_{d\times d})$, then at least one intrinsic mean exists and since $(\mathbb{P}_{d\times d}, g_R)$ is a geodesically complete manifold of non-positive curvature, the intrinsic mean μ is also unique. By (Pennec, 2006, Corollary 1), the intrinsic mean is conveniently represented by $\mu \in \mathbb{P}_{d\times d}$ satisfying,

$$\boldsymbol{E}_\nu[\mathrm{Log}_\mu(X)] \;=\; \mathbf{0}.$$

Here, $\mathbf{0}$ is the zero matrix and $\boldsymbol{E}_\nu[\cdot]$ is the Euclidean mean in the space of Hermitian matrices. The sample intrinsic mean typically has no closed-form solution, but it can be computed efficiently through gradient descent as detailed in e.g., Pennec (2006).

In the remainder of this section, $\gamma : \mathcal{I} \to \mathbb{P}_{d\times d}$, with $\mathcal{I} \subset \mathbb{R}$, is assumed to be a square integrable matrix-valued curve, such that $\int_{\mathcal{I}} \delta_R(\gamma(u), y_0)^2\, du < \infty$ for some $y_0 \in \mathbb{P}_{d\times d}$. As input data observations we consider a finite sequence of intrinsic local averages $M_{J,k} = \mathrm{Ave}_{I_{J,k}}(\gamma)$, across equally-sized non-overlapping intervals $(I_{J,k})_k$ with $0 \le k \le n-1$, such that $\bigcup_k I_{J,k} = \mathcal{I}$. Here, $\mathrm{Ave}_{I_{J,k}}(\gamma)$ denotes the intrinsic mean of γ over the interval $I_{J,k}$. Without loss of generality, it is assumed that $\mathcal{I} = [0,1]$ and that $n = 2^J$ is dyadic in order to allow for a straightforward construction of the scaling coefficient pyramid below. The latter is not an absolute limitation of the approach, as the intrinsic wavelet transforms can also be adapted to non-dyadic observation grids, as explained in more detail in (Chau, 2018, Chapter 5).

2.1 Intrinsic AI refinement scheme

Midpoint pyramid The construction of the wavelet transforms is based on the idea of *lifting* transforms. For an overview of first- and second-generation wavelet transforms using the lifting scheme, we refer to e.g., Jansen and Oonincx (2005) or Klees and Haagmans (2000). First, we build a redundant midpoint or scaling coefficient pyramid analogous to Rahman et al. (2005), starting with the sequence of midpoint coefficients $(M_{J,k})_k$ at the finest scale J. At the next coarser scale $j = J-1$ set,

$$M_{j,k} \;:=\; \eta(M_{j+1,2k}, M_{j+1,2k+1}, 1/2), \quad \text{for } k = 0,\ldots,2^j-1, \tag{2.1}$$

and continue this coarsening operation up to scale $j = 0$, such that each scale j contains a total of 2^j midpoints. Here, $\eta(p_1,p_2,1/2)$ is the halfway point or *midpoint* on the geodesic

segment connecting $p_1, p_2 \in \mathbb{P}_{d\times d}$, which coincides with the intrinsic sample mean of p_1 and p_2. As a convenient way to write an intrinsic weighted or unweighted sample mean, we also use the notation $\mathrm{Ave}(\cdot;\cdot)$. That is, if $X_1, \ldots, X_n \in \mathbb{P}_{d\times d}$, then $\overline{X}_n = \mathrm{Ave}(\{X_i\}_i; \{w_i\}_i)$ is the weighted intrinsic average of $X_1, \ldots, X_n$ with weights $w_1, \ldots, w_n$, such that $\overline{X}_n$ solves:

$$\overline{X}_n \;=\; \mathrm{Exp}_{\overline{X}_n}\left(\sum_{i=1}^{n} w_i \mathrm{Log}_{\overline{X}_n}(X_i)\right). \tag{2.2}$$

If we write $\overline{X}_n = \mathrm{Ave}(\{X_i\}_i)$, then $\overline{X}_n$ is understood to be the unweighted intrinsic average of $X_1, \ldots, X_n$. In particular, we can write in a recursive fashion $M_{j,k} := \mathrm{Ave}(\{M_{j+1,2k}, M_{j+1,2k+1}\})$.

Intrinsic polynomials Intrinsic polynomials as defined in Hinkle et al. (2014) play a key role in the construction of the AI refinement scheme. Essentially, polynomial curves of degree $k \geq 0$ in the Riemannian manifold are defined as the curves with vanishing k-th and higher order covariant derivatives. Let $\gamma : \mathcal{I} \to \mathbb{P}_{d\times d}$ be a smooth curve on the manifold, with existing covariant derivatives of all orders, then it is said to be a polynomial curve of degree k if,

$$\nabla^{\ell}_{\gamma'}\gamma'(t) \;:=\; (\nabla_{\gamma'})^{\ell}\gamma'(t) \;=\; \mathbf{0}, \qquad \forall\, \ell \geq k \text{ and } t \in \mathcal{I},$$

where $\nabla^0_{\gamma'}\gamma'(t) := \gamma'(t)$. A zero degree polynomial is a curve for which $\gamma'(t) = \mathbf{0}$, i.e., a constant curve. A first-degree polynomial is a curve for which $\nabla_{\gamma'}\gamma'(t) = \mathbf{0}$ corresponding to a geodesic curve, i.e., a *straight* line in the manifold. In general, higher degree polynomials are difficult to represent in closed form, but discretized polynomial curves are straightforward to generate via numerical integration as described in Hinkle et al. (2014).

Intrinsic polynomial interpolation At scale $j \in \{0, \ldots, J-1\}$, the intrinsic AI refinement scheme takes as input coarse-scale midpoints $(M_{j,k})_k$ and outputs imputed or predicted finer-scale midpoints $(\widetilde{M}_{j+1,k'})_{k'}$. The predicted midpoints are computed as the $(j+1)$-scale midpoints of the unique intrinsic polynomial $\tilde{\gamma} : \mathcal{I} \to \mathbb{P}_{d\times d}$ with j-scale midpoints $(M_{j,k})_k$. In order to reconstruct intrinsic polynomials from a discrete set of points on the manifold, we consider a generalized intrinsic version of Neville's algorithm as in (Ma and Fu, 2012, Chapter 9.2), replacing ordinary linear interpolation by *geodesic interpolation*.

Given $P_0, \ldots, P_n \in \mathbb{P}_{d\times d}$ and $x_0 < \ldots < x_n \in \mathbb{R}$, set $p_{i,i}(x) := P_i$ for all x and $i = 0, \ldots, n$. The $p_{i,i}$ are zero-th order polynomials, since $p'_{i,i}(x) = \mathbf{0}$. Iteratively define,

$$p_{i,j}(x) \;:=\; \mathrm{Exp}_{p_{i,j-1}(x)}\left(\frac{x - x_i}{x_j - x_i}\mathrm{Log}_{p_{i,j-1}(x)}(p_{i+1,j}(x))\right), \qquad 0 \leq i < j \leq n,$$

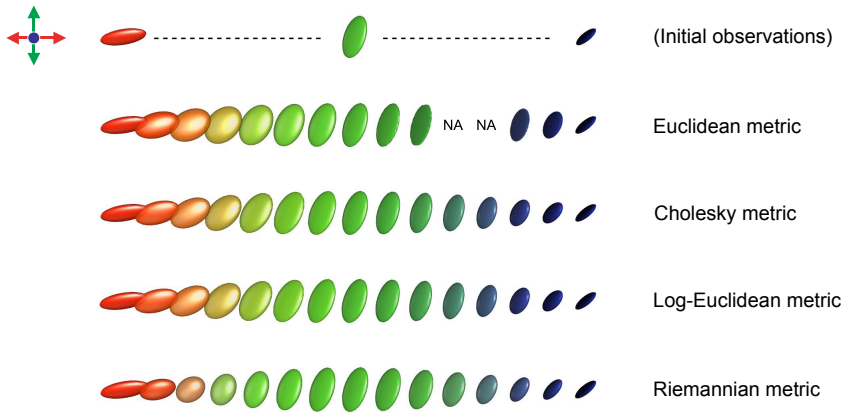

Figure 1: Illustration of intrinsic polynomial interpolation for (3×3)-SPD matrices represented as 3D-ellipsoids. The colors indicate the direction of the principal eigenvectors.

where $p_{i+1,j}(x)$ and $p_{i,j-1}(x)$ are the intrinsic polynomials of degree $j-i-1$ passing through $P_{i+1},\dots,P_j$ at $x_{i+1},\dots,x_j$ and through $P_i,\dots,P_{j-1}$ at $x_i,\dots,x_{j-1}$ respectively. Then $p_{i,j}(x)$ is the intrinsic polynomial of degree $j-i$ passing through $P_i,\dots,P_j$ at $x_i,\dots,x_j$. Continuing the above iterative reconstruction, at the final iteration we obtain the intrinsic polynomial $p_{0,n}(x)$ of order n passing through $P_0,\dots,P_n$ at $x_0,\dots,x_n$.

To illustrate, $p_{0,1}(x)$ is the geodesic, i.e., first-order intrinsic polynomial, passing through P_0 and P_1 at x_0 and x_1. In general, since $p_{i,j}(x)$ geodesically interpolates two polynomials of degree $j-i-1$, $p_{i,j}(x)$ is itself a polynomial of degree $j-i$ introducing one additional higher-degree non-vanishing covariant derivative. This is exactly analogous to the Euclidean setting, where linear interpolation of two polynomials of degree r results in a polynomial of degree at most $r+1$. Intrinsic polynomial interpolation for a curve of HPD matrices by means of Neville's algorithm is implemented through the function `pdNeville()` in the `pdSpecEst`-package, which is demonstrated in Figure 1 by the interpolation of three (3×3)-dimensional SPD matrices represented as 3D-ellipsoids with respect to a number of different metrics on the space of HPD matrices. Note that the interpolating second-order polynomial subject to the Euclidean metric is not everywhere positive definite, as indicated by the `NA` values.

2.1.1 Midpoint prediction via intrinsic average interpolation

Reconstructing the intrinsic polynomial $\tilde{\gamma}(x)$ with j-scale midpoints $(M_{j,k})_k$ is not equivalent to reconstructing the intrinsic polynomial passing through the j-scale midpoints, which corre-

sponds to an *interpolating* refinement scheme instead of an *average*-interpolating refinement scheme. Interpolating wavelet transforms are not well-suited to noise-removal applications, as noise would not get averaged out at coarser scales in the associated scaling coefficient pyramid. This is also discussed in more detail in Rahman et al. (2005) and (Chau, 2018, Chapter 2). To compute predicted midpoints via intrinsic average-interpolating refinement, instead we consider the cumulative intrinsic mean of $\tilde{\gamma}(x)$, $M_{y_0} : (y_0, 1] \to \mathbb{P}_{d\times d}$, given by:

$$M_{y_0}(y) \quad = \quad \mathrm{Ave}_{[y_0,y]}(\tilde{\gamma}). \tag{2.3}$$

If $\tilde{\gamma}(x)$ is the intrinsic polynomial with j-scale midpoints $(M_{j,k})_{k=0,\ldots,2^j-1}$, then $M_0((k+1)2^{-j})$ equals the cumulative intrinsic average of $\{M_{j,0}, \ldots, M_{j,k-1}\}$. The key consideration is that the cumulative intrinsic mean of an intrinsic polynomial of order r is again an intrinsic polynomial of order $\leq r$. For instance, given a geodesic segment, i.e., a first-order polynomial, its cumulative intrinsic mean is a geodesic segment moving at half the original speed. Again, this is analogous to the Euclidean setting, where an integrated polynomial remains a polynomial.

Fix a location $k \in \{L, \ldots, 2^{(j-1)} - (L+1)\}$ at scale $j-1$ for some $L \geq 0$. Given the neighboring $(j-1)$-scale midpoints $\{M_{j-1,k-L}, \ldots, M_{j-1,k}, \ldots, M_{j-1,k+L}\}$, then we predict the finer-scale midpoints $\{M_{j,2k}, M_{j,2k+1}\}$. Here, $N := 2L + 1 \geq 0$ is referred to as the order or degree of the refinement scheme. First, to predict the midpoint $M_{j,2k+1}$, fit an intrinsic polynomial $\widehat{M}_{(k-L)2^{-(j-1)}}(y)$ of order $N-1$ through the N known points $\{\overline{M}_{j-1,0}, \ldots, \overline{M}_{j-1,N-1}\}$ by means of Neville's algorithm, where $\overline{M}_{j-1,\ell}$ denotes the cumulative intrinsic average:

$$\overline{M}_{j-1,\ell} \quad := \quad M_{(k-L)2^{-(j-1)}}((k-L+\ell)2^{-(j-1)}) \quad = \quad \mathrm{Ave}(\{M_{j-1,i}\}_{i=k-L}^{k-L+\ell}). \tag{2.4}$$

By construction of the cumulative intrinsic mean curve, $M_{(k-L)2^{-(j-1)}}((2k+1)2^{-j})$ lies on the geodesic segment connecting the known cumulative average $\overline{M}_{j-1,L}$ and the midpoint $M_{j,2k+1}$. Replacing $M_{(k-L)2^{-(j-1)}}((2k+1)2^{-j})$ by its estimate $\widehat{M}_{(k-L)2^{-(j-1)}}((2k+1)2^{-j})$, the following expression for the predicted midpoint $\widetilde{M}_{j,2k+1}$ can be derived:

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

For the exact derivation, see the proof of Proposition 3.2 in the supplementary material. The value of $\widetilde{M}_{j,2k}$ directly follows from the midpoint relation $\mathrm{Ave}(\{\widetilde{M}_{j,2k}, \widetilde{M}_{j,2k+1}\}) = M_{j-1,k}$ as,

$$\widetilde{M}_{j,2k} \quad = \quad M_{j-1,k} * \widetilde{M}_{j,2k+1}^{-1}.$$

An important observation is that if the coarse-scale midpoints $\{M_{j-1,k-L}, \ldots, M_{j-1,k+L}\}$ are generated from an intrinsic polynomial $\gamma(x)$ of degree $\leq N-1$, then the midpoints $\{M_{j,2k}, M_{j,2k+1}\}$ are reproduced without error. This is analogous to the scalar AI refinement scheme in Donoho (1993) and is referred to as the *intrinsic polynomial reproduction* property.

If $k \in \{0, \ldots, L-1\} \cup \{2^{(j-1)} - (L-1), \ldots, 2^{(j-1)} - 1\}$ is located near the boundary, not all symmetric neighbors around $M_{j-1,k}$ are available for prediction of $\{M_{j,2k}, M_{j,2k+1}\}$. Instead, collect the N closest neighbors of $M_{j-1,k}$ either to the left or right and predict the j-scale midpoints as above through $(N-1)$-th order intrinsic polynomial interpolation based on the non-symmetric neighbors $(M_{j-1,k+\ell})_\ell$. This boundary modification preserves the intrinsic polynomial reproduction property.

2.1.2 Faster midpoint prediction in practice

In the scalar AI refinement scheme on the real line in Donoho (1993) or (Klees and Haagmans, 2000, pg. 95), the predicted j-scale scaling coefficients obtained via polynomial average-interpolation of the $(j-1)$-scale scaling coefficients are equivalent to weighted linear combinations of the input scaling coefficients, with weights depending on the average-interpolation order N. In the intrinsic version of Neville's algorithm, the only change with respect to its Euclidean counterpart is the nature of the interpolation, i.e., linear interpolation is substituted by geodesic interpolation. The predicted midpoints $\{\widetilde{M}_{j,2k}, \widetilde{M}_{j,2k+1}\}$ remain weighted averages of the inputs $\{M_{j-1,k-L}, \ldots, M_{j-1,k}, \ldots, M_{j-1,k+L}\}$, with the same weights as in the Euclidean case. The difference being that –instead of weighted Euclidean averages– the weighted averages are intrinsic weighted averages in the Riemannian manifold. That is,

$$\begin{aligned} \widetilde{M}_{j,2k} &= \text{Ave}\left(\{M_{j-1,k+\ell}\}_{\ell=-L}^{L}; \{C_{N,2\ell+N-1}\}_{\ell=-L}^{L}\right) \\ \widetilde{M}_{j,2k+1} &= \text{Ave}\left(\{M_{j-1,k+\ell}\}_{\ell=-L}^{L}; \{C_{N,2\ell+N}\}_{\ell=-L}^{L}\right), \end{aligned} \tag{2.5}$$

where the weights $\boldsymbol{C}_N = (C_{N,i})_{i=0,\ldots,2N-1}$ depend on the refinement order $N \geq 1$ and sum up to 2. For instance, away from the boundary; for $N=1$, $\boldsymbol{C}_1 = (1,1)$; for $N=3$, $\boldsymbol{C}_3 = (1,-1,8,8,-1,1)/8$; for $N=5$, $\boldsymbol{C}_5 = (-3,3,22,-22,128,128,-22,22,3,-3)/128$; and for $N=7$, $\boldsymbol{C}_7 = (5,-5,-44,44,201,-201,1024,1024,-201,201,44,-44,-5,5)/1024$. In the `pdSpecEst`-package, these prediction weights are pre-determined up to order $N \leq 9$ at locations *away of* and *at* the boundary, allowing for faster computation of the predicted midpoints in practice. For $N > 9$, the midpoints are predicted via the intrinsic version of Neville's algorithm.

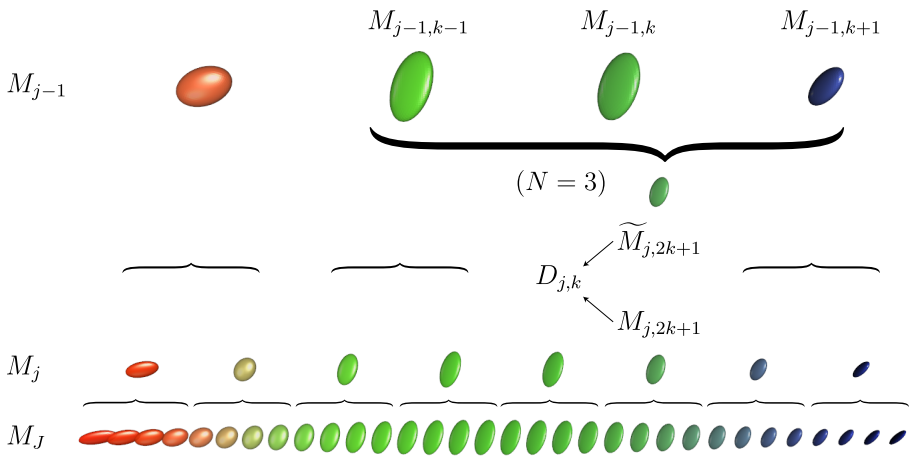

Figure 2: Illustration of intrinsic foward AI wavelet transform for a *smooth* curve of three (3×3)-dimensional SPD matrices represented as 3D-ellipsoids at different midpoint scales.

2.2 Intrinsic forward and backward AI wavelet transform

Forward wavelet transform The intrinsic AI refinement scheme leads to an intrinsic AI wavelet transform passing from j-scale midpoints to $(j-1)$-scale midpoints plus j-scale wavelet coefficients. The steps in the intrinsic AI wavelet transform are also visualized in Figure 2 based on a sequence of (3×3)-dimensional SPD matrices represented as 3D-ellipsoids.

1. **Coarsen/Predict:** given j-scale midpoints $(M_{j,k})_{k=0,\ldots,2^j-1}$, compute the $(j-1)$-scale midpoints $(M_{j-1,k})_{k=0,\ldots,2^{j-1}-1}$ via the midpoint relation in eq.(2.1). Select a refinement order $N \geq 1$ and generate the predicted midpoints $(\widetilde{M}_{j,k})_{k=0,\ldots,2^j-1}$ based on $(M_{j-1,k})_k$.

2. **Difference:** given the true and predicted j-scale midpoints $M_{j,2k+1}, \widetilde{M}_{j,2k+1}$, define the wavelet coefficients as an *intrinsic difference* according to,

$$D_{j,k} \;=\; 2^{-j/2}\mathrm{Log}_{\widetilde{M}_{j,2k+1}}\big(M_{j,2k+1}\big) \;\in\; T_{\widetilde{M}_{j,2k+1}}(\mathbb{P}_{d\times d}). \tag{2.6}$$

   Note that $\|D_{j,k}\|^2_{\widetilde{M}_{j,2k+1}} = 2^{-j}\delta_R(M_{j,2k+1}, \widetilde{M}_{j,2k+1})^2$ by definition of the Riemannian distance, giving the wavelet coefficients the interpretation of a (scaled) difference between $M_{j,2k+1}$ and $\widetilde{M}_{j,2k+1}$. In addition, we also keep track of the *whitened* wavelet coefficients,

$$\mathfrak{D}_{j,k} \;=\; \widetilde{M}^{-1/2}_{j,2k+1} * D_{j,k} \;\in\; T_{\mathrm{Id}}(\mathbb{P}_{d\times d}), \tag{2.7}$$

   with $\mathrm{Id} \in \mathbb{P}_{d\times d}$ the $(d \times d)$-dimensional identity matrix. The whitened coefficients correspond to the coefficients in eq.(2.6) transported to the same tangent space (at the

identity). This allows for straightforward comparison of coefficients across scales and locations in Section 3 and 4, since $\|\mathfrak{D}_{j,k}\|_F^2 = \|D_{j,k}\|^2_{\widetilde{M}_{j,2k+1}}$.

Backward wavelet transform The backward wavelet transform passing from coarse $(j-1)$-scale midpoints plus j-scale wavelet coefficients to finer j-scale midpoints follows from reverting the above operations:

1. **Predict/Refine**: given $(j-1)$-scale midpoints $(M_{j-1,k})_{k=0,\ldots,2^{j-1}-1}$ and a refinement order $N \geq 1$, generate the predicted midpoints $(\widetilde{M}_{j,2k+1})_{k=0,\ldots,2^{j-1}-1}$ and compute the j-scale midpoints at the odd locations $2k+1$ for $k = 0, \ldots, 2^j - 1$ through:

$$M_{j,2k+1} \quad = \quad \mathrm{Exp}_{\widetilde{M}_{j,2k+1}}(2^{j/2} D_{j,k}).$$

2. **Complete**: the j-scale midpoints at the even locations $2k$ for $k = 0, \ldots, 2^j - 1$ are retrieved from $M_{j-1,k}$ and $M_{j,2k+1}$ through the midpoint relation in eq.(2.1) as,

$$M_{j,2k} \quad = \quad M_{j-1,k} * M_{j,2k+1}^{-1}.$$

Given the coarsest midpoint $M_{0,0}$ at scale $j = 0$ and the wavelet coefficient pyramid $(D_{j,k})_{j,k}$, for $j = 1, \ldots, J$ and $k = 0, \ldots, 2^{j-1} - 1$, repeating the reconstruction procedure above up to scale J, we retrieve the original input sequence of local averages $(M_{J,k})_k$ for $k = 0, \ldots, 2^J - 1$.

3 Wavelet regression for smooth HPD curves

In this section, we derive the wavelet coefficient decay and linear wavelet thresholding convergence rates in the context of the intrinsic AI wavelet transforms for intrinsically smooth curves of HPD matrices. It turns out that the derived rates coincide with the usual scalar wavelet coefficient decay and linear thresholding convergence rates on the real line. Nonlinear thresholding will not improve the convergence rates in the case of a homogeneous smoothness space. However, nonlinear wavelet thresholding is expected to improve the convergence rates in the case of globally non-homogeneous smoothness spaces. This requires a well-defined intrinsic generalization to the Riemannian manifold of e.g., the family of Besov smoothness spaces, which is outside the scope of this paper.

Repeated midpoint operator The repeated midpoint operator in eq.(2.1) in the construction of the midpoint pyramid is a valid intrinsic averaging operator in the sense that it converges to the intrinsic mean in the metric space $(\mathbb{P}_{d\times d}, \delta_R)$ at a rate conjectured in Rahman et al. (2005) for general Riemannian manifolds. As in Rahman et al. (2005), recursively define,

$$\mu_n \;:=\; \mu_n(X_1,\ldots,X_n) \;=\; \mathrm{Ave}\big(\{\mu_{n/2}(X_1,\ldots,X_{n/2}), \mu_{n/2}(X_{n/2+1},\ldots,X_n)\}\big). \tag{3.1}$$

Proposition 3.1. *(Convergence midpoint operator) Let* $X_1,\ldots,X_n \overset{\text{iid}}{\sim} \nu$*, such that* $\nu \in P_2(\mathbb{P}_{d\times d})$ *with intrinsic mean* $\mathbb{E}_\nu[X] = \mu$*, and* $n = 2^J$ *for some* $J > 0$*. Then,*

$$\boldsymbol{E}[\delta_R(\mu_n,\mu)^2] \;\lesssim\; n^{-1},$$

with $\lesssim$ *smaller or equal up to a constant. Moreover,* $\mu_n \xrightarrow{p} \mu$ *as* $n\to\infty$*, where the convergence holds with respect to the Riemannian distance, i.e., for every* $\epsilon > 0$*,* $P(\delta_R(\mu_n,\mu) > \epsilon) \to 0$*.*

Wavelet coefficient decay of smooth curves The derivation of the wavelet coefficient decay of intrinsically smooth curves in the Riemannian manifold relies on the fact that the derivative $\gamma'(t) \in T_{\gamma(t)}(\mathbb{P}_{d\times d})$ of a smooth curve $\gamma : \mathcal{I} \to \mathbb{P}_{d\times d}$ can be Taylor expanded in terms of the parallel transport and covariant derivatives according to (Lang, 1995, Chapter 9, Proposition 5.1) as,

$$\gamma'(t) \;=\; \sum_{k=0}^{m} \Gamma(\gamma)_{t_0}^{t}\left(\nabla_{\gamma'}^k \gamma'(t_0)\right)\frac{(t-t_0)^k}{k!} + O((t-t_0)^{m+1}), \qquad \text{as } t\to t_0, \tag{3.2}$$

where the parallel transport $\Gamma(\gamma)_{t_0}^{t}(v)$ transports a vector $v \in T_{\gamma(t_0)}(\mathbb{P}_{d\times d})$ to the tangent space $T_{\gamma(t)}(\mathbb{P}_{d\times d})$ along the curve γ. If $\gamma(t)$ is an intrinsic polynomial curve of order $r > 0$, then, since $\Gamma(\gamma)_{t_0}^{t}(\mathbf{0}) = \mathbf{0}$, all terms of order higher or equal to r vanish and $\gamma'(t)$ simplifies to,

$$\gamma'(t) \;=\; \sum_{k=0}^{r-1} \Gamma(\gamma)_{t_0}^{t}(\nabla_{\gamma'}^k \gamma'(t_0))\frac{(t-t_0)^k}{k!}.$$

In the specific case of a first-order polynomial, the above expression reduces to $\gamma'(t) = \Gamma(\gamma)_{t_0}^{t}(\gamma'(t_0))$, i.e., γ' is parallel transported along the curve γ itself, or in other words, $\gamma(t)$ is a geodesic curve.

Proposition 3.2. *(Coefficient decay) Given a refinement order* $N \geq 1$*, suppose that* $\gamma : [0,1] \to \mathbb{P}_{d\times d}$ *is a smooth curve with existing covariant derivatives of order* N *or higher. Then, for each scale* $j > 0$ *sufficiently large and location* k*,*

$$\|\mathfrak{D}_{j,k}\|_F \;\lesssim\; 2^{-j/2}2^{-jN},$$

where $\mathfrak{D}_{j,k}$ denotes the whitened wavelet coefficient at scale-location (j,k) as in eq.(2.7) obtained from the intrinsic AI wavelet transform with refinement order N. Here, the finest-scale midpoints are given by the local intrinsic averages $M_{J,k} = \mathrm{Ave}_{I_{J,k}}(\gamma)$, with $I_{J,k} = [k/2^J, (k+1)/2^J]$ for $k = 0, \ldots, 2^J - 1$.

Remark Note that the above decay rates correspond to the usual wavelet coefficient decay rates of smooth real-valued curves in a Euclidean space based on wavelets with N vanishing moments, see e.g., (Walnut, 2002, Theorem 9.5).

Consistency and convergence rates The following results detail the convergence rates of linear thresholding of wavelet scales of intrinsically smooth curves $\gamma : [0,1] \to \mathbb{P}_{d\times d}$ corrupted by noise. Let $M_{J,k} = \mathrm{Ave}_{I_{J,k}}(\gamma)$, with $I_{J,k} = [k/n, (k+1)/n]$ for $k = 0, \ldots, n-1$ as before, and suppose that $X_0, \ldots, X_{n-1}$ is an independent sample, such that $X_k \sim \nu_k$ with $\nu_k \in P_2(\mathbb{P}_{d\times d})$ and $\mathbb{E}_{\nu_k}[X] = M_{J,k}$ for each $k = 0, \ldots, n-1$. The proposition below gives the estimation error of the empirical wavelet coefficients based on $X_0, \ldots, X_{n-1}$ with respect to the true wavelet coefficients based on the sequence $M_{J,0}, \ldots, M_{J,n-1}$. The proof relies on the convergence rate in Proposition 3.1 above.

Proposition 3.3. *(Estimation error) Let $M_{J,0}, \ldots, M_{J,n-1}$ and $X_0, \ldots, X_{n-1}$ be as defined above, with $n = 2^J$ for some $J > 0$. Then, for each scale $j > 0$ sufficiently small and each location k, it holds that,*

$$\boldsymbol{E}\|\widehat{\mathfrak{D}}_{j,k,n} - \mathfrak{D}_{j,k}\|_F^2 \quad \lesssim \quad n^{-1},$$

where $\widehat{\mathfrak{D}}_{j,k,n} = 2^{-j/2}\,\mathrm{Log}(\widetilde{M}_{j,2k+1,n}^{-1/2} * M_{j,2k+1,n})$ is the empirical whitened wavelet coefficient at scale-location (j,k), with $M_{j,2k+1,n}$ the estimated repeated midpoint at scale-location $(j, 2k+1)$ based on $X_0, \ldots, X_{n-1}$ and $\widetilde{M}_{j,2k+1,n}$ the predicted midpoint based on the estimated midpoints $(M_{j-1,k',n})_{k'}$ and some refinement order $N \geq 1$.

Combining Proposition 3.2 and 3.3, the main theorem below provides the averaged mean squared Riemannian error of a linear wavelet estimator of a smooth curve $\gamma(t)$ based on the sample of observations $X_0, \ldots, X_{n-1}$. Again, the convergence rates correspond to the usual nonparametric convergence rates of linear wavelet estimators of smooth real-valued curves in a Euclidean space based on wavelets with N vanishing moments, see e.g., Antoniadis (1997).

Theorem 3.4. *(Convergence rates linear thresholding) Given a refinement order $N \geq 1$, suppose that $\gamma : [0,1] \to \mathbb{P}_{d\times d}$ is a smooth curve with existing covariant derivatives of order N or higher, and let $M_{J,0}, \ldots, M_{J,n-1}$ and $X_0, \ldots, X_{n-1}$ be as defined above, with $n = 2^J$ for some $J \geq 0$. Consider the linear wavelet estimator based on the observations $X_0, \ldots, X_{n-1}$ that thresholds all wavelet coefficients at scales $j \geq J_0$, such that $J_0 = \log_2(n)/(2N+1)$, with N the order of the intrinsic AI wavelet transform. For n sufficiently large,*

$$\sum_{j,k} \boldsymbol{E}\|\widehat{\mathfrak{D}}_{j,k,n} - \mathfrak{D}_{j,k}\|_F^2 \quad \lesssim \quad n^{-2N/(2N+1)}, \tag{3.3}$$

where $\widehat{\mathfrak{D}}_{j,k,n}$ is the empirical whitened wavelet coefficient after linear thresholding of wavelet scales and the sum ranges over all scales $1 \leq j \leq J$ and locations $0 \leq k \leq 2^{j-1}-1$. Moreover, denote by $(\widehat{M}_{J,k,n})_k$ the finest-scale midpoints based on the linear thresholded wavelet estimator. Then, for n sufficiently large, also,

$$\frac{1}{n}\sum_{k=0}^{n-1} \boldsymbol{E}\left[\delta_R\big(M_{J,k}, \widehat{M}_{J,k,n}\big)^2\right] \quad \lesssim \quad n^{-2N/(2N+1)}. \tag{3.4}$$

Remark Denoting $\hat{\gamma}_n(t) = \widehat{M}_{J,k,n}\mathbf{1}_{\{t\in I_{J,k}\}}$ and $\gamma_n(t) = M_{J,k}\mathbf{1}_{\{t\in I_{J,k}\}}$, with $\mathbf{1}$ the indicator function. If it is further assumed that $\gamma(t) - \gamma_n(t) = O(n^{-N/(2N+1)})$ for $t \in [0,1]$, then the linear wavelet estimator $\hat{\gamma}_n(t)$ converges to the continuous curve γ at the same rate as in Theorem 3.4 above,

$$\int_0^1 \boldsymbol{E}\left[\delta_R(\hat{\gamma}_n(t), \gamma(t))^2\right] dt \quad \lesssim \quad n^{-2N/(2N+1)}.$$

The derivation of this result follows directly from the application of a generalized triangle inequality, the details of which can be found in Appendix II in the supplementary material.

4 Wavelet-based spectral matrix estimation

In the context of multivariate spectral matrix estimation, consider data observations from a d-dimensional strictly stationary time series of length $T = 2n$ with HPD spectral density matrix $f(\omega) \in \mathbb{P}_{d\times d}$ and raw periodogram matrix $I_T(\omega_\ell)$ at the Fourier frequencies $\omega_\ell = \pi\ell/n \in (0,\pi]$ for $\ell = 1, \ldots, n$. The aim of this section is to estimate $f(\omega)$ by denoising the inconsistent spectral estimator $I_T(\omega_\ell)$ through shrinkage or thresholding of coefficients in the intrinsic wavelet domain. Given the setup in Section 2.1, we can define the equally-sized intervals $I_{J,k} = (\pi k/n, \pi(k+1)/n]$, with $0 \leq k \leq n-1$ and $\bigcup_k I_{J,k} = (0,\pi]$, such that each interval $I_{J,k}$

contains a single Fourier frequency ω_{k+1}. As we only consider estimating the spectrum at the Fourier frequencies, we set the finest-scale local averages to $M_{J,k} = \text{Ave}_{I_{J,k}}(f) = f(\omega_{k+1})$.

Pre-smoothed periodogram By construction, the raw periodogram matrix $I_T(\omega_\ell)$ is Hermitian, but only positive semidefinite as the rank of $I_T(\omega_\ell)$ is one. The intrinsic wavelet transform acts on curves of HPD matrices and for this reason we pre-smooth the periodogram to guarantee that it is HPD or full rank analogous to e.g., Dai and Guo (2004). By (Dai and Guo, 2004, Lemma 1), for $\omega_\ell \not\equiv 0 \pmod{\pi}$, a multitaper spectral estimate $\bar{I}_T(\omega_\ell)$ of a strictly stationary time series, with a fixed number of tapers L, is asymptotically independent at the Fourier frequencies, and its asymptotic distribution satisfies:

$$\bar{I}_T(\omega_\ell) \quad \stackrel{d}{\to} \quad W_d^C(L, L^{-1}f(\omega_\ell)), \qquad \text{as } n \to \infty.$$

Here, $W_d^C(L, L^{-1}f(\omega_\ell))$ denotes a complex Wishart distribution of dimension d with L degrees of freedom and Euclidean mean $f(\omega_\ell)$. If $L \geq d$, then the spectral estimate $\bar{I}_T(\omega_\ell)$ is positive definite with probability one. In order to pre-smooth the raw periodogram matrix $I_T(\omega_\ell)$, we choose $L = d$ as small as possible, so that only the necessary small amount of pre-smoothing is performed to guarantee an HPD periodogram matrix $\bar{I}_T(\omega_\ell)$.

Asymptotic bias-correction Suppose that $X \sim W_d^c(L, L^{-1}f)$ exactly, then the Euclidean mean of X equals f, and if $X_1, \ldots, X_n \overset{\text{iid}}{\sim} W_d^c(L, L^{-1}f)$, the ordinary arithmetic mean $\frac{1}{n}\sum_{\ell=1}^n X_\ell$ is an unbiased and consistent estimator of f as $n \to \infty$. Intrinsic averaging in the midpoint pyramid is performed through repeated application of the midpoint operator. By Proposition 3.1, it is understood that if the Euclidean mean $\boldsymbol{E}[X_\ell] = f$ and the intrinsic mean $\mathbb{E}[X_\ell] = \mu$ do not coincide, the repeated midpoint functional is not a consistent estimator of f, the quantity of interest. By defining the notion of intrinsic bias as in Smith (2000), the repeated midpoint functional of a multitaper spectral estimate is seen to be asymptotically biased with respect to the spectrum f.

Definition 4.1. Given an estimator $\hat{\mu}$ of $\mu \in \mathbb{P}_{d\times d}$, define the bias $b(\hat{\mu}, \mu) \in T_\mu(\mathbb{P}_{d\times d})$ of $\hat{\mu}$ as,

$$b(\hat{\mu}, \mu) \quad = \quad \boldsymbol{E}[\text{Log}_\mu(\hat{\mu})].$$

Note that in a Euclidean space, the Exp- and Log-maps reduce to ordinary matrix addition and subtraction. In this case the above definition is seen to simplify to the usual vector space definition of the bias.

Theorem 4.1. *(Bias-correction) Let* $X \sim W_d^c(L, L^{-1}f)$ *and* $c(d,L) = -\log(L) + \frac{1}{d}\sum_{i=1}^{d}\psi(L-(d-i))$*, with* $\psi(\cdot)$ *the digamma function. The intrinsic bias of* X *in relation to* f *is given by,*

$$b(X,f) \;=\; \boldsymbol{E}[\mathrm{Log}_f(X)] \;=\; c(d,L)\cdot f.$$

If $(\widetilde{X}_\ell)_{\ell=1,\ldots,n} := (e^{-c(d,L)}X_\ell)_{\ell=1,\ldots,n}$*, such that* $X_1,\ldots,X_n \overset{\text{iid}}{\sim} W_d^c(L,L^{-1}f)$ *with* $n = 2^J$*, then,*

$$\mu_n(\widetilde{X}_1,\ldots,\widetilde{X}_n) \;\xrightarrow{p}\; f, \qquad \text{as } n\to\infty,$$

where the convergence in probability holds with respect to the Riemannian distance.

Remark It is pointed out that if $d = L = 1$, the bias-correction simplifies to multiplication by the scalar $\exp(-c(d,L)) = \exp(-\psi(1))$, the exponential of the Euler-Mascheroni constant. This corresponds to the asymptotic bias-correction for the ordinary log-periodogram with respect to the log-spectrum in the context of a univariate time series, see e.g., Wahba (1980).

Remark As the bias-corrected (and pre-smoothed) periodogram $\bar{I}_T(\omega_\ell)$ is asymptotically equivalent in distribution to a sequence of bias-corrected independent Wishart matrices, linear wavelet estimation of the bias-corrected periodogram approximately enjoys the same convergence properties as discussed at the end of Section 3.

4.1 Nonlinear intrinsic wavelet thresholding

Given a sequence of d-dimensional time series observations, wavelet-based spectral estimation exploits the sparsity of representations of smooth curves in the intrinsic AI wavelet domain by proceeding along the usual steps:

1. Apply the intrinsic AI wavelet transform to the bias-corrected HPD periodogram.
2. Shrink or threshold the coefficients in the intrinsic wavelet domain.
3. Apply the inverse intrinsic AI wavelet transform to the modified coefficients.

There are various possible ways to nonlinearly shrink or threshold coefficients in the intrinsic manifold wavelet domain. In particular, expanding the matrix-valued coefficients in a basis of the vector space of Hermitian matrices, nonlinear thresholding or shrinkage of individual components allows to capture inhomogeneous smoothness behavior across components of the spectral matrix, similar to the Cholesky-based smoothing procedures in e.g., Dai and Guo (2004) or Krafty and Collinge (2013). The wavelet-denoised estimator is guaranteed to

be HPD, as the inverse wavelet transform always outputs a curve in the manifold of HPD matrices. From the perspective of wavelet coefficients being intrinsic local differences in the manifold, another sensible approach is to shrink or threshold all components of a matrix-valued wavelet coefficient simultaneously, e.g., a kink or cusp in a curve in the manifold likely affects all components of the matrix-valued wavelet coefficients at the corresponding scale-locations instead of a single or only a few components. Here, we pursue the latter approach and consider keep-or-kill thresholding of entire wavelet coefficients.

Congruence equivariance In general, the only requirement that is imposed on the intrinsic wavelet thresholding or shrinkage procedure is that it is *unitary congruence equivariant*. That is, if D^X is a noisy matrix-valued wavelet coefficient and $\widehat{D}^X$ is its shrunken or thresholded version, then $U * \widehat{D}^X$ should be the shrunken or thresholded version of $U * D^X$ for each $U \in \mathcal{U}_d$, where $\mathcal{U}_d = \{U \in \mathrm{GL}(d, \mathbb{C}) \mid U^*U = \mathrm{Id}\}$, and $\mathrm{GL}(d, \mathbb{C}) = \{A \in \mathbb{C}^{d\times d} \mid \det(A) \neq 0\}$ denotes the general linear group of $d \times d$ invertible complex matrices. In practice, this property virtually always holds. For instance, if one thresholds or shrinks components of coefficients data-adaptively, the component-specific threshold or shrinkage parameters rotate in the same fashion as the components of the coefficients.

Proposition 4.2. *(Unitary congruence equivariance) Let $(X_\ell)_\ell$ be a sequence of HPD matrices and $(\hat{f}_\ell)_\ell$ its wavelet-denoised estimate. If the wavelet thresholding or shrinkage procedure is unitary congruence equivariant, then the same is true for the wavelet estimator, i.e., the wavelet-denoised estimate of $(U * X_\ell)_\ell$ is $(U * \hat{f}_\ell)_\ell$ for each $U \in \mathcal{U}_d$.*

This is an important property in the context of multivariate spectral estimation. Rotation of the observed time series data, e.g., permuting the time series components, results in a congruence transformation $U * f(\omega)$ of the generating spectral matrix, with $U \in \mathcal{U}_d$. Such rotations should not nontrivially affect the spectral estimator, as the observed rotation of the time series is essentially an arbitrary representation of the data. The spectral estimation methods based on smoothing the Cholesky decomposition of an initial noisy spectral estimator (Dai and Guo (2004), Rosen and Stoffer (2007) or Krafty and Collinge (2013)) do not necessarily satisfy this condition. This is due to the fact that Cholesky square root matrices are generally not unitary congruence-equivariant, i.e., $\mathrm{Chol}(U * f(\omega)) \neq U * \mathrm{Chol}(f(\omega))$ for a non-trivial unitary matrix $U \in \mathcal{U}_d$. To circumvent this problem, in Zheng et al. (2017), the authors propose to average a large set of Cholesky-based estimates based on random rotations of the data. The main

drawback of such an approach is the significant increase in computational effort.

Trace thresholding of coefficients A method that is particularly *traceable* is thresholding or shrinkage based on the trace of the whitened wavelet coefficients. For a sequence of independent complex Wishart matrices, the trace of the noisy whitened coefficients decomposes into an additive signal plus mean-zero noise sequence model. Moreover, the variance of the trace of the noisy whitened coefficients is constant across wavelet scales, and since the trace operator outputs a scalar, one can directly apply ordinary scalar thresholding or shrinkage methods to the matrix-valued coefficients. Thresholding or shrinkage of the trace of the whitened coefficients is equivariant under unitary congruence transformations as in Proposition 4.2. Moreover, it is equivariant under congruence transformation by any invertible matrix, i.e., *general linear* congruence equivariant. In the context of spectral estimation of multivariate time series, this means that the estimator does not nontrivially depend on the chosen basis or coordinate system of the time series, as the spectral estimator is equivariant under a change of basis of the time series.

Lemma 4.3. *(General linear congruence equivariance) Let $(X_\ell)_\ell$ be a sequence of HPD matrices and $(\hat{f}_\ell)_\ell$ its wavelet-denoised estimate based on linear or nonlinear shrinkage of the trace of the whitened wavelet coefficients. The estimator is equivariant under general linear congruence transformation in the sense that the wavelet-denoised estimate $(\hat{f}_{A,\ell})_\ell$ of $(A * X_\ell)_\ell$ equals $(A * \hat{f}_\ell)_\ell$ for each $A \in \mathrm{GL}(d, \mathbb{C})$.*

In the following, $\widetilde{P}_f$ denotes the probability distribution associated to a bias-corrected complex Wishart distribution $e^{-c(d,L)} W_d^c(L, L^{-1} f)$ as in Theorem 4.1, with $L \geq d$ to ensure positive-definiteness of the Wishart matrix. Here, $\widetilde{P}_f \in P_2(\mathbb{P}_{d\times d})$ is understood to be the distribution of a random variable $X = f^{1/2} * W$, where W is an HPD complex Wishart matrix, with L degrees of freedom, not depending on f, and with intrinsic mean equal to the identity matrix Id. Note that the latter directly implies that the intrinsic mean of $f^{1/2} * W$ is equal to f.

Proposition 4.4. *(Trace properties) Let $X_\ell \sim \widetilde{P}_{f_\ell}$, independently distributed for $\ell = 1, \ldots, n$, with $n = 2^J$. For each scale-location (j, k), the whitened wavelet coefficients obtained from the intrinsic AI wavelet transform of order $N = 2L + 1 \geq 1$ satisfy:*

$$\mathrm{Tr}(\mathfrak{D}^X_{j,k}) \quad = \quad \mathrm{Tr}(\mathfrak{D}^f_{j,k}) + \mathrm{Tr}(\mathfrak{D}^W_{j,k}),$$

where $\mathfrak{D}^X_{j,k}$ *is the random whitened coefficient based on the sequence* $(X_\ell)_{\ell=1}^n$, $\mathfrak{D}^f_{j,k}$ *is the deterministic whitened coefficient based on the sequence of intrinsic means* $(f_\ell)_{\ell=1}^n$, *and* $\mathfrak{D}^W_{j,k}$ *is the random whitened coefficient based on an i.i.d. sequence of Wishart matrices* $(W_\ell)_{\ell=1}^n$, *with intrinsic mean equal to the identity, independent of* $(f_\ell)_{\ell=1}^n$.

Moreover, $\boldsymbol{E}[\mathrm{Tr}(\mathfrak{D}^X_{j,k})] = \mathrm{Tr}(\mathfrak{D}^f_{j,k})$, *and,*

$$\mathrm{Var}(\mathrm{Tr}(\mathfrak{D}^X_{j,k})) = \left(2^{-J}\sum_{i=0}^{2N-1} \boldsymbol{C}^2_{L,i}\right)\left(\sum_{i=1}^{d}\psi'(L-(d-i))\right), \tag{4.1}$$

where $\psi'(\cdot)$ *denotes the trigamma function, and* $(\boldsymbol{C}_{L,i})_i$ *are the filter coefficients as in eq.(2.5). In particular,* $\mathrm{Var}(\mathrm{Tr}(\mathfrak{D}^X_{j,k}))$ *is independent of the scale-location* (j,k) *and whenever* $\mathrm{Tr}(\mathfrak{D}^f_{j,k})$ *vanishes* $\boldsymbol{E}[\mathrm{Tr}(\mathfrak{D}^X_{j,k})] = 0$, *e.g., when* $(f_\ell)_\ell$ *is sampled from an intrinsic polynomial of order smaller than* N.

Corollary 4.5. *(Centered noise) With the same notation as in Proposition 4.4, the random whitened wavelet coefficients* $\mathfrak{D}^W_{j,k}$ *based on a sequence of i.i.d. Wishart matrices* $(W_\ell)_{\ell=1}^n$, *with identity intrinsic mean satisfy,*

$$\boldsymbol{E}[\mathfrak{D}^W_{j,k}] = \mathbf{0},$$

where $\boldsymbol{E}[\cdot]$ *denotes the (ordinary) Euclidean expectation.*

Based on the trace of the whitened coefficients, by Proposition 4.4, in the context of a sequence of approximate complex random Wishart matrices, such as a curve of periodogram matrices, any preferred standard wavelet shrinkage procedure can be applied well-suited to scalar additive signal plus noise sequence models, with homogeneous variances across coefficient scales.

5 Illustrative data examples

5.1 Finite-sample performance

Simulation setup In the figures below, we assess the finite-sample performance of intrinsic wavelet-based curve estimation in the space of HPD matrices and benchmark the performance against several alternative nonparametric smoothing procedures. In particular, we consider HPD test curves displaying both globally homogeneous and locally varying smoothness behavior, available through the function `rExamples1D()` in the `pdSpecEst`-package. The `arma` spectrum is a smooth (2×2)-dimensional HPD spectral matrix generated by a stationary

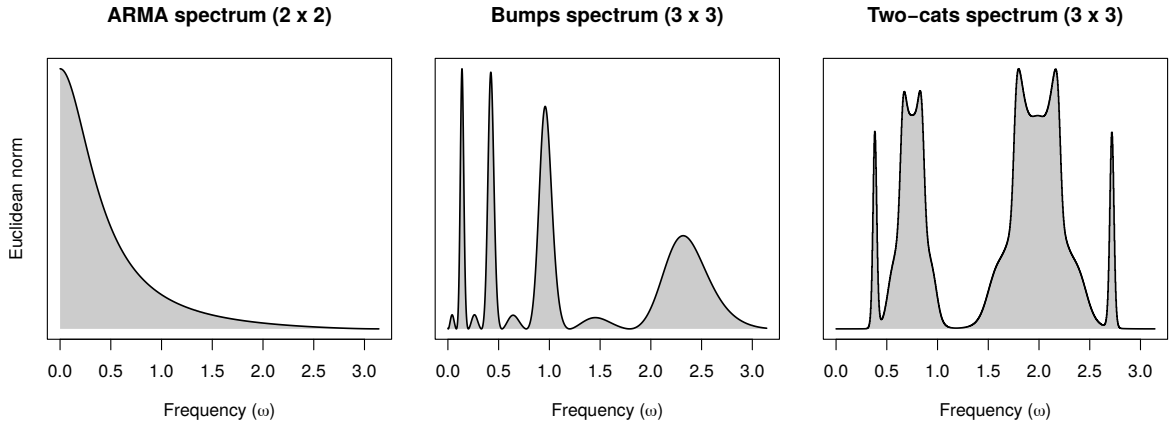

Figure 3: Euclidean norms of HPD test spectral matrices generated with `rExamples1D()`.

ARMA$(1,1)$ process based on (Brockwell and Davis, 2006, Example 11.4.1). The `bumps` spectrum is a curve of (3×3)-dimensional HPD matrices containing local bumps of various degrees of smoothness, and the `two-cats` spectrum visualizes the contours of two cats and consists of relatively smooth parts combined with local peaks and troughs. Figure 3 displays the Euclidean norm of the HPD matrix-valued curves as a function of frequency. Each test spectrum is normalized to have unit Euclidean norm over the integrated frequency range $[0,\pi]$.

Given the HPD test curves, random observations $(X_\ell)_{\ell=1,\ldots,n} \in \mathbb{P}_{d\times d}$ are generated according to several different model distributions centered around the target curve $(f_\ell)_{\ell=1,\ldots,n} \in \mathbb{P}_{d\times d}$. The data generating models and associated metrics used for estimation are summarized in Table 2. In the periodogram noise scenario, first a d-dimensional time series trace is generated from the target spectrum f via its Cramér representation with complex normal random variates as in e.g., (Brillinger, 1981, Section 4.6), and second an initial HPD multitaper periodogram $(X_\ell)_\ell$ is computed based on d discrete prolate spheroidal (DPSS) taper functions. The observations $(X_\ell)_\ell$ tend in distribution to the Wishart noise scenario as the length of the time series increases. The scale of the noise distributions in the Log-Gaussian and Riemannian Gaussian noise scenarios is chosen such that the signal-to-noise ratio is comparable to the Wishart and periodogram noise scenarios. Additional details on the intrinsic signal-noise model in the Riemannian-Gaussian noise scenario are found in (Chau, 2018, Section 2.2.6).

In each individual simulation experiment, the intrinsic integrated squared estimation error (IISE) is calculated as the integrated squared error based on the distance associated to the metric used for estimation. These are, respectively, the Riemannian distance δ_R; the Log-

Table 2: Simulation setup and signal-noise models.

Simulation scenario	Signal-noise model	Noise distribution*	Metric	B-C†
Wishart noise	$X_\ell = f_\ell^{1/2} * Z_\ell$	$Z_\ell \overset{\text{iid}}{\sim} \frac{1}{d} W_d^C(d, \mathrm{Id})$	Riemannian Cholesky	✓ ✓
Log-Gaussian noise	$X_\ell = \mathrm{Exp}(\mathrm{Log}(f_\ell) + Z_\ell)$	$Z_\ell \overset{d}{=} \sum_{k=1}^{d^2} z_k e^k$, $(z_k)_k \overset{\text{iid}}{\sim} N(0, 1/4)$	Log-Euclidean	✗
Riem.-Gaussian noise	$X_\ell = f_\ell^{1/2} * Z_\ell$	$Z_\ell \overset{d}{=} \sum_{k=1}^{d^2} z_k e^k$, $(z_k)_k \overset{\text{iid}}{\sim} N(0, 1/4)$	Riemannian	✗
Periodogram noise	$X_\ell = \bar{I}_T(\omega_\ell)$	Multitaper periodogram with d DPSS tapers	Riemannian	✓

∗: $\{e^1, \ldots, e^{d^2}\} \in \mathbb{H}_{d\times d}$ is an orthonormal basis of $(\mathbb{H}_{d\times d}, \langle\cdot,\cdot\rangle_F)$.

†: B-C denotes whether a bias-correction is required for the given data generating scenario and metric.

Euclidean distance $\delta_L(x, y) = \|\mathrm{Log}(y) - \mathrm{Log}(x)\|_F$; and the Cholesky distance $\delta_C(x, y) = \|\mathrm{Chol}(y) - \mathrm{Chol}(x)\|_F$, with $x, y \in \mathbb{P}_{d\times d}$. For the Wishart noise scenario, HPD matrix curve estimation subject to the Riemannian or the Cholesky metric is biased with respect to the target HPD matrix curve. Under the Cholesky metric, this bias can be corrected by the bias-correction in (Dai and Guo, 2004, Theorem 1). Under the Riemannian metric, we apply the bias-correction in Theorem 4.1. For the periodogram noise scenario, we again make use of the bias-correction in Theorem 4.1. For the Log-Gaussian noise scenario and estimation subject to the Log-Euclidean metric, the estimators are unbiased and no bias-correction is necessary. The same holds true for the Riemannian-Gaussian noise scenario and estimation with respect to the Riemannian metric.

Estimation procedures The simulation experiments include linear thresholding of wavelet scales according to Section 3 and nonlinear trace thresholding of wavelet coefficients as in Section 4.1 in the space of HPD matrices equipped with the Riemannian, Log-Euclidean or Cholesky metric. As a straightforward nonlinear thresholding method, we consider scalar dyadic tree-structured thresholding based on the wavelet coefficient traces similar to Donoho (1997). More precisely, for each scale-location (j, k), denote $d_{j,k} = \mathrm{Tr}(\mathfrak{D}_{j,k}^X)$ for the trace of the observed whitened wavelet coefficient and let $w_{j,k} \in \{0, 1\}$ be a binary label. Given a

regularization parameter $\lambda \geq 0$, we optimize the following complexity penalized loss criterion:

$$\underset{\boldsymbol{w}}{\operatorname{argmin}}\, L(\boldsymbol{w}) \quad = \quad \underset{\boldsymbol{w}}{\operatorname{argmin}} \sum_{j,k} |d_{j,k} w_{j,k} - d_{j,k}|^2 + \lambda^2 \sum_{j,k} w_{j,k}, \tag{5.1}$$

under the constraint that the nonzero labels $\{w_{j,k} \,|\, w_{j,k} = 1\}$ form a dyadic rooted tree, i.e., for each nonzero label $w_{j+1,2k+1}$ or $w_{j+1,2k}$, the label $w_{j,k}$ also has to be nonzero. This minimization problem can be solved in $O(n)$ computations via the tree-pruning algorithm in Donoho (1997), with n the total number of coefficients, resulting in the estimated wavelet coefficients $\widehat{D}_{j,k} = w_{j,k} D^X_{j,k}$. Linear and nonlinear tree-structured wavelet thresholding are available in the `pdSpecEst`-package through the function `pdSpecEst1D()` and the argument `metric` set to the appropriate metric. The choice `metric = "Riemannian-Rahman"` replaces the forward and backward AI wavelet transforms by the MI wavelet transforms in Rahman et al. (2005) based on the affine-invariant Riemannian metric. The latter is slightly different to the metric suggested in (Rahman et al., 2005, Section 4.4), which does not enjoy the same congruence invariance properties as the Riemannian metric. In addition to intrinsic wavelet-based curve estimation, we have implemented intrinsic versions of the following curve estimation procedures in the space of HPD matrices equipped with the Riemannian, Log-Euclidean and Cholesky metric: (i) Nearest-Neighbor (NN) regression, (ii) Cubic Spline (CS) regression (as in Boumal and Absil (2011b), Boumal and Absil (2011a)) and (iii) Local Polynomial (LP) regression (as in Yuan et al. (2012)). In the periodogram noise scenario, a benchmark multitaper spectral estimator based on the generated time series has also been included. Details about the listed estimation procedures are found in Appendix III in the supplementary material.

Simulation results Figure 4 displays boxplots of the (relative) IISE distributions, sampled at $n = 256$ and $n = 512$ locations, based on $M = 10\ 000$ replications per scenario. For each simulation scenario, the obtained IISEs are standardized with respect to the IISE of the linear wavelet estimator, in order to allow for easier comparison between metrics and noise scenarios. That is, all displayed outputs above the horizontal dashed unit line perform worse than linear wavelet thresholding, and all outputs below the horizontal line outperform linear wavelet thresholding. Each estimation procedure depends on a single main tuning parameter: for the linear wavelet regression, this is the number of nonzero wavelet scales; for the tree-structured wavelet and cubic spline regression, this is the regularization parameter; for the nearest neighbor regression, this is the number of nearest neighbors; for the local polynomial regression, this is the bandwidth parameter; and for the multitaper estimation procedure,

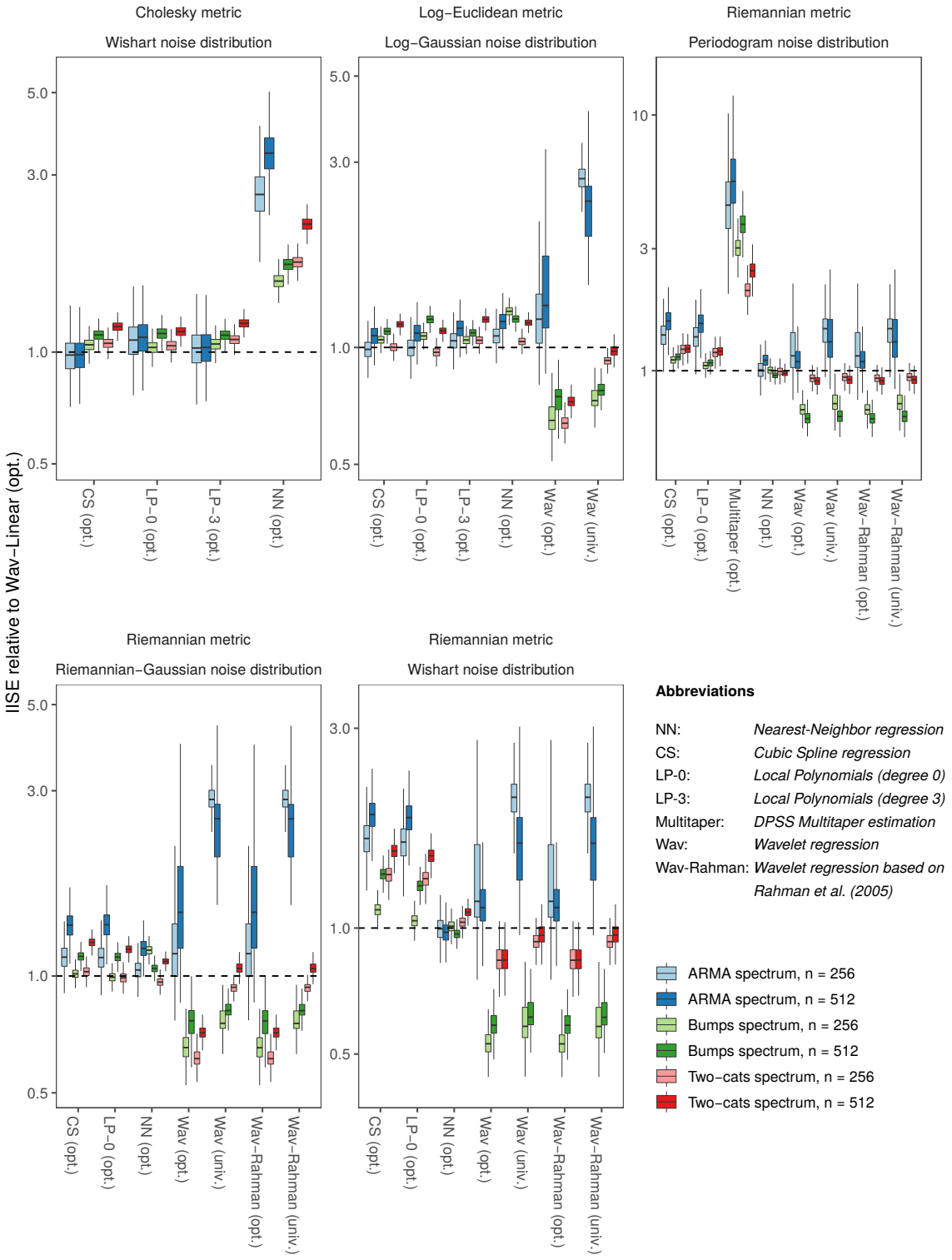

Figure 4: Relative intrinsic integrated squared estimation errors (IISE) of the wavelet and benchmark estimation procedures for the `arma`, `bumps` and `two-cats` test spectra, relative to the IISE of intrinsic linear wavelet thresholding.

this is the number of tapering functions. In each simulation experiment, an oracle tuning parameter, denoted by `(opt.)`, is determined by minimizing the IISE with respect to the true target HPD matrix curve. In addition, for the tree-structure wavelet estimators a choice of the regularization parameter based on an ordinary universal threshold is included, denoted by `(univ.)`. For the Wishart noise scenario subject to the Cholesky metric, nonlinear wavelet thresholding has been excluded from the simulation experiments, as the traces of the wavelet coefficients cannot be shown to decompose into a scalar additive signal plus noise sequence model as in Proposition 4.4, which is the case for the other simulation scenarios based on the Riemannian and Log-Euclidean metric.

Intrinsic nonlinear wavelet thresholding outperforms linear wavelet thresholding in terms of the IISE for the `bumps` spectrum and to a somewhat lesser extent for the `two-cats` spectrum. This is attributed to the fact that, in contrast to the benchmark procedures and linear wavelet thresholding, the nonlinear wavelet estimator is able to capture varying degrees of smoothness in the HPD matrix curve. On the other hand, the benchmark procedures and linear wavelet thresholding do outperform nonlinear wavelet thresholding in terms of the IISE in the highly smooth `arma` spectrum, as a single global smoothing parameter is sufficient to capture the smooth behavior in the HPD spectral matrix. Furthermore, –in the majority of the simulation scenarios– the IISE of nonlinear tree-structured thresholding based on a simple universal threshold is relatively close to the optimal IISE for nonlinear tree-structured thresholding, thereby providing a fast heuristic choice of the main tuning parameter in practical applications. For all considered benchmark procedures, there is no simple heuristic choice for the main tuning parameter(s), and one needs to resort either to computationally expensive cross-validation methods or manual hyperparameter tuning.

5.2 Associative learning experiment LFP data

To further demonstrate the appeal of the intrinsic wavelet methods, we consider spectral matrix estimation for a subset of brain signal time series trials recorded over the course of an associative learning experiment with a macaque, see Gorrostieta et al. (2012) or Fiecas and Ombao (2016) for additional details. During the learning experiment, the electrical activity in the brain of the macaque is measured by means of local field potentials (LFP). After preprocessing of the LFP time series, there remain a total of $S = 590$ trial-specific approximately stationary 2-dimensional time series traces of length $T = 2048$ sampled at $1\,000\,\text{Hz}$. The

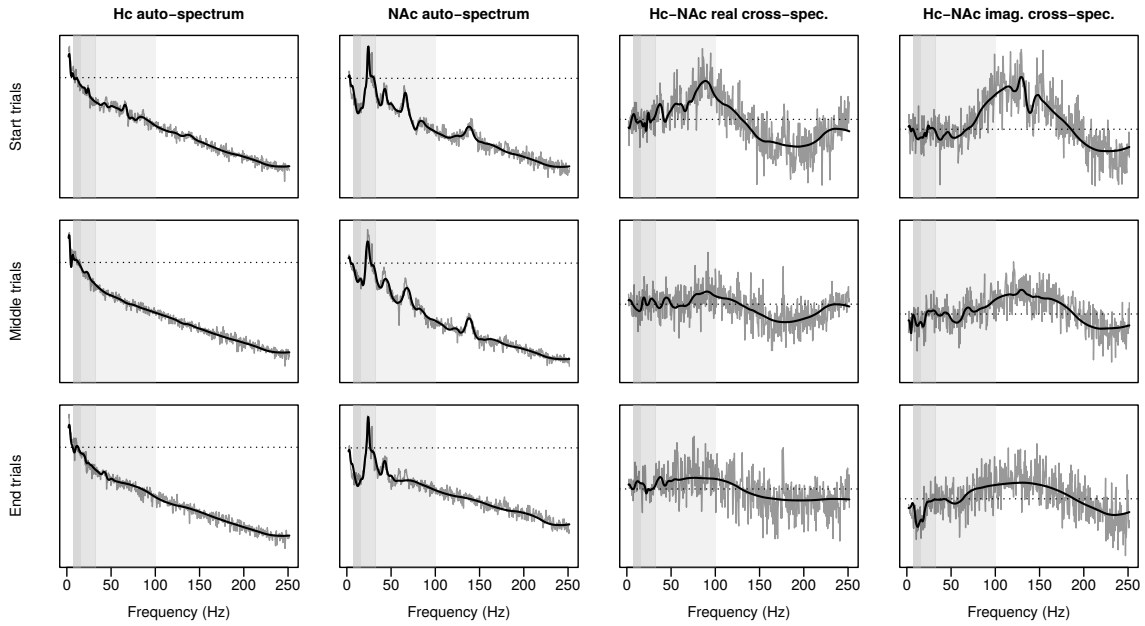

Figure 5: Row-expanded matrix logarithms of HPD periodograms averaged across LFP time series trials at the start, middle and end of the learning experiment across Fourier frequencies.

two time series components correspond to LFP measurements in the hippocampus (Hc) and nucleus accumbens (NAc) regions of the macaque's brain, which have previously been implicated in cognitive processes involving memory and reward, as detailed in Fiecas and Ombao (2016) and the references therein. For demonstrational purposes, we extract trials from the start of the experiment ($s = 1, \ldots, 10$), the middle of the experiment ($s = 291, \ldots, 300$) and the end of the experiment ($s = 581, \ldots, 590$). For each of the trial subsets, an averaged HPD (2×2)-periodogram matrix is computed by averaging the trial-specific raw (2×2)-periodogram matrices across Fourier frequencies. Figure 5 displays the matrix logarithms of the initial noisy HPD periodograms up to 250 Hz averaged across LFP trial subsets. The grey bands display respectively the α-band (8-16 Hz), the β-band (16-32 Hz) and the γ-band (32-100 Hz). The overlayed black lines correspond to nonlinear wavelet denoised HPD periodograms subject to the Riemannian metric obtained with the function `pdSpecEst1D()`, with refinement order $N = 5$ and tree-structured trace thresholding based on a rescaled universal threshold.

Let $f(\omega)$ denote the theoretical (2×2)-dimensional HPD spectral matrix of the stationary LFP time series process at frequency ω. Among other steps, preprocessing of the raw LFP time series data includes standardizing the time series traces to have zero mean and unit variance. After standardization, the spectral matrix transforms as $A * f(\omega)$, with A given by

some diagonal matrix $A = ((\theta_1, 0)', (0, \theta_2)')$. If one also permutes the order of the time series traces, the matrix A becomes $A = ((0, \theta_1)', (\theta_2, 0)')$. These are two straightforward examples of preprocessing steps that ideally should not have a nontrivial impact on the final spectral estimator as previously argued in Section 4.1. In Figure 6, we demonstrate the effects such transformations can have on the estimation of the LFP spectral matrices, focusing on the periodogram data associated to the middle of the experiment. Here, $M = 1\,000$ random matrices $A_m \in \mathrm{GL}(2, \mathbb{C})$ with standard complex Gaussian matrix entries are generated. First, the initial HPD periodograms $I_T(\omega)$ are transformed by $A_m * I_T(\omega)$, imitating a basis transformation of the LFP time series data. Second, a linear wavelet thresholded spectrum $\hat{f}_m(\omega)$ is calculated, discarding all coefficients above scale $J = 4$. Denoising by means of linear thresholding allows for straightforward visual comparisons between different metrics. Third, the spectral estimates are transformed back to the original basis of the LFP time series data, according to $A_m^{-1} * \hat{f}_m(\omega)$. Under the Cholesky metric, the same procedure is repeated with random unitary matrices $A_m \in \mathcal{U}_2$, sampled with respect to the (additively invariant) Haar measure on $\mathcal{U}_2$. The black lines in Figure 6 display the spectral estimate $\hat{f}(\omega)$ based on the original periodogram data. The grey regions include all spectral estimates $\hat{f}_m(\omega)$ subject to congruence transformation by the random matrices A_m. For the Log-Euclidean and Cholesky metric, the estimated Hc and NAc auto-spectral components are nearly equivariant, but the estimated cross-spectral components potentially display a high degree of non-equivariance depending on the choice of A_m. Note that this observed non-equivariance directly extends to the estimated coherence, which are obtained as normalized versions of the estimated cross-spectra.

6 Concluding remarks

The primary contribution of this paper is the development of intrinsic average-interpolation (AI) wavelet transforms and intrinsic wavelet thresholding for curves in the space of HPD matrices equipped with the affine-invariant Riemannian metric. The intrinsic wavelet transforms are constructed independent of the chosen metric and although the wavelet coefficient decay and nonparametric convergence rates in Section 3 are derived exclusively for the affine-invariant Riemannian metric, similar arguments apply to other metrics as well. For instance, in a Euclidean space the intrinsic Taylor expansions reduce to ordinary Taylor expansions, as the parallel transport is the identity map and the covariant derivatives are standard ma-

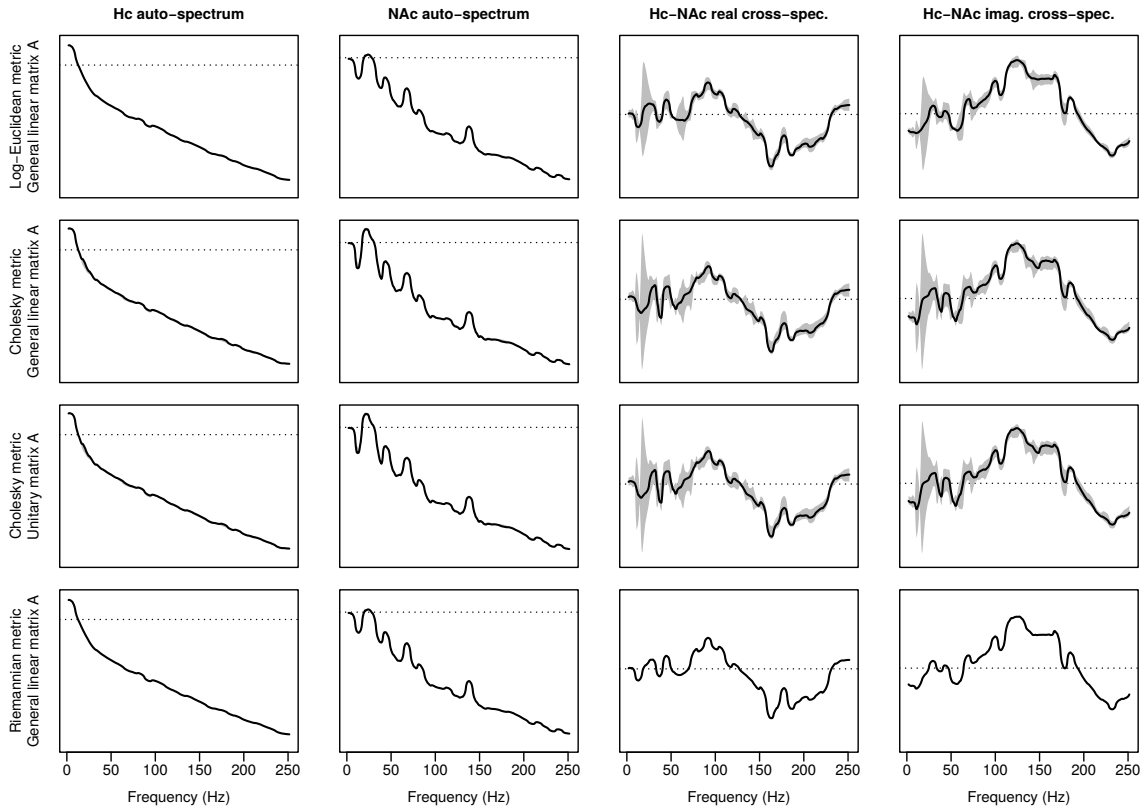

Figure 6: Row-expanded matrix logarithms of linear wavelet thresholded spectral estimates $\hat{f}(\omega)$ and $\hat{f}_m(\omega)$ subject to the Log-Euclidean, Cholesky and Riemannian metrics.

trix derivatives. In the context of spectral matrix estimation of high-dimensional time series, spectral estimation with respect to the Riemannian metric may suffer from computational instability, as the estimation target may be close to or at the boundary of the space. Besides the Euclidean metric, alternative metrics that can handle rank deficient spectral matrices include e.g., the Procrustes shape-and-size metric or the Cholesky metric, see Dryden et al. (2009). However, polynomial interpolation with respect to the Procrustes metric may lead to negative definite matrices, similar to the Euclidean metric, and the Cholesky square root matrix is not necessarily unique in the rank deficient case. The challenge of flexible estimation of nonnegative definite spectral matrices is currently a topic of interest for future research. Furthermore, Hermitian or symmetric positive definite matrices are encountered as autocovariance matrices or spectral density matrices in time series analysis, but also play an important role in the fields of medical imaging, computer vision or radar signal processing (e.g., Pennec et al. (2006)), and it is of interest to apply the intrinsic wavelet methods for the purpose of compression or de-

noising in other settings than spectral matrix estimation. For instance, applied to diffusion tensor imaging, intrinsic wavelet shrinkage or thresholding shows potential for fast denoising of large collections of non-smoothly varying diffusion tensors.

Additional material In Chau et al. (2019), the notion of intrinsic data depth in the space of HPD matrices equipped with the affine-invariant Riemannian metric is discussed, providing a center-to-outward ordering of a collection of HPD matrices. In the context of HPD spectral matrix estimation, the data depths are useful tools to construct confidence regions for the spectral matrix –intrinsic to the Riemannian geometry of the space– based on for instance a parametric bootstrap using the data generating process of a stationary time series via its Cramér representation as detailed in Dai and Guo (2004) and Fiecas and Ombao (2016) among others. In (Chau, 2018, Chapter 5), the intrinsic wavelet methods presented in this paper are extended to *surfaces* of Hermitian positive definite matrices, with in mind the application to nonparametric estimation of the time-varying spectrum of a locally stationary time series. In addition to spectral matrix estimation, the intrinsic wavelet transforms are also useful for e.g., fast clustering of spectral matrices based on sparse representations in the intrinsic wavelet domain. Implementations of these methods are available in the `pdSpecEst`-package, and we refer to the package documentation (Chau (2017)) for further descriptions.

Acknowledgements

The authors gratefully acknowledge financial support from the following agencies and projects: the Belgian Fund for Scientific Research FRIA/FRS-FNRS (J. Chau), the contract "Projet d'Actions de Recherche Concertées" No. 12/17-045 of the "Communauté française de Belgique" (R. von Sachs), IAP research network P7/06 of the Belgian government (R. von Sachs). We thank Hernando Ombao and Dr. Emad Eskandar (Massachussetts General Hospital) for providing access to the local field potential data.

References

Antoniadis, A. (1997). Wavelets in statistics: a review. *Statistical methods & applications 6*(2), 97–130.

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.

Brillinger, D. (1981). *Time Series: Data Analysis and Theory*. San Francisco: Holden-Day.

Brockwell, P. and R. Davis (2006). *Time Series: Theory and Methods*. New York: Springer.

Chau, J. (2017). *pdSpecEst: An Analysis Toolbox for Hermitian Positive Definite Matrices*. R Package version 1.2.3.

Chau, J. (2018). *Advances in Spectral Analysis for Multivariate, Nonstationary and Replicated Time Series*. Ph. D. thesis, Université catholique de Louvain.

Chau, J., H. Ombao, and R. von Sachs (2019). Intrinsic data depth for Hermitian positive definite matrices. *Journal of Computational and Graphical Statistics*. (To appear).

Dahlhaus, R. (2012). *Locally stationary processes*, Chapter in Time Series Analysis: Methods and Applications, Vol. 30, pp. 351–413. Amsterdam: Elsevier.

Dai, M. and W. Guo (2004). Multivariate spectral analysis using Cholesky decomposition. *Biometrika 91*(3), 629–643.

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

Donoho, D. (1993). *Smooth wavelet decompositions with blocky coefficient kernels*, Chapter in Recent Advances in Wavelet Analysis, pp. 259–308. New York: Academic Press.

Donoho, D. (1997). Cart and best-ortho-basis: a connection. *The Annals of Statistics 25*(5), 1870–1911.

Dryden, I., A. Koloydenko, and D. Zhou (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics 3*(3), 1102–1123.

Fiecas, M. and H. Ombao (2016). Modeling the evolution of dynamic brain processes during an associative learning experiment. *Journal of the American Statistical Association 111*(516), 1440–1453.

Gorrostieta, C., H. Ombao, R. Prado, S. Patel, and E. Eskandar (2012). Exploring dependence between brain signals in a monkey during learning. *Journal of Time Series Analysis 33*(5), 771–778.

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

Hinkle, J., P. Fletcher, and S. Joshi (2014). Intrinsic polynomials for regression on Riemannian

manifolds. *Journal of Mathematical Imaging and Vision 50*(1-2), 32–52.

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

Holbrook, A., S. Lan, A. Vandenberg-Rodes, and B. Shahbaba (2018). Geodesic Lagrangian Monte Carlo over the space of positive definite matrices: with application to Bayesian spectral density estimation. *Journal of Statistical Computation and Simulation 88*(5), 982–1002.

Jansen, M. and P. Oonincx (2005). *Second Generation Wavelets and Applications*. London: Springer-Verlag.

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.

Klees, R. and R. Haagmans (2000). *Wavelets in the Geosciences*. Berlin: Springer-Verlag.

Krafty, R. and W. Collinge (2013). Penalized multivariate Whittle likelihood for power spectrum estimation. *Biometrika 100*(2), 447–458.

Lang, S. (1995). *Differential and Riemannian Manifolds*. New York: Springer-Verlag.

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

Ma, Y. and Y. Fu (2012). *Manifold Learning Theory and Applications*. CRC Press, Taylor & Francis.

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

Pasternak, O., N. Sochen, and P. Basser (2010). The effect of metric selection on the analysis of diffusion tensor MRI data. *NeuroImage 49*(3), 2190–2204.

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.

Rahman, I., I. Drori, V. Stodden, D. Donoho, and P. Schröder (2005). Multiscale representations for manifold-valued data. *Multiscale Modeling & Simulation 4*(4), 1201–1232.

Rosen, O. and D. Stoffer (2007). Automatic estimation of multivariate spectra via smoothing splines. *Biometrika 94*(2), 335–345.

Said, S., L. Bombrun, Y. Berthoumieu, and J. Manton (2017). Riemannian Gaussian distributions on the space of symmetric positive definite matrices. *IEEE Transactions on Information Theory 63*(4), 2153–2170.

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.

Wahba, G. (1980). Automatic smoothing of the log periodogram. *Journal of the American Statistical Association 75*(369), 122–132.

Walden, A. (2000). A unified view of multitaper multivariate spectral estimation. *Biometrika 87*(4), 767–788.

Walnut, D. (2002). *An Introduction to Wavelet Analysis*. Boston: Birkhäuser.

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.

Zheng, H., K.-W. Tsui, X. Kang, and X. Deng (2017). Cholesky-based model averaging for covariance matrix estimation. *Statistical Theory and Related Fields 1*(1), 48–58.

Zhu, H., Y. Chen, J. Ibrahim, Y. Li, C. Hall, and W. Lin (2009). Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *Journal of the American Statistical Association 104*(487), 1203–1212.

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 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 $a \in \mathrm{GL}(d, \mathbb{C})$, 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,

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

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,

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

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,

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

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 $\mathrm{Log}_x(y) = y - x$ 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:

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

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). 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).

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:

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

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 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:

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

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

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

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

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

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

As a consequence, also,

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

□

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:

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

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

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

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,

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

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), 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,

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

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 $j = 1$,

$$\begin{aligned}
\widehat{M}_{1,k,n} &= \mathrm{Exp}_{\widehat{\widetilde{M}}_{1,k,n}}\big(2^{1/2}\widehat{D}_{1,k,n}\big) \\
&= \widetilde{M}_{1,k,n}^{1/2} * \mathrm{Exp}\big(2^{1/2}\widetilde{M}_{1,k,n}^{-1/2} * \widehat{D}_{1,k,n}\big) \\
&= \widetilde{M}_{1,k,n}^{1/2} * \mathrm{Exp}\big(2^{1/2}\widehat{\mathfrak{D}}_{1,k,n}\big) \\
&= \Big(\widetilde{M}_{1,k} + \epsilon_{1,n}\Big)^{1/2} * \mathrm{Exp}\Big(2^{1/2}(\mathfrak{D}_{1,k} + \eta_n)\Big) \\
&= \Big(\widetilde{M}_{1,k}^{1/2} + \epsilon_{1,n}\Big) * \Big(\mathrm{Exp}(2^{1/2}\mathfrak{D}_{1,k}) + 2^{1/2}\eta_n\Big) \\
&= M_{1,k} + O_p(2^{1/2}n^{-1/2}) + O_p(2^{-(J-1)/2}) \\
&= M_{1,k} + O_p(2^{1/2}n^{-1/2}). \qquad (8.8)
\end{aligned}$$

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 $j = J_0$,

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

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:

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

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. □

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,

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

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), 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:

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

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 $g(\mathrm{Tr}(\mathfrak{D}^X_{j,k})) \in \mathbb{R}$ be a thresholding or

shrinkage rule depending on $\mathrm{Tr}(\mathfrak{D}^X_{j,k})$, such that $D^{\hat{f}}_{j,k} = g(\mathrm{Tr}(\mathfrak{D}^X_{j,k}))D^X_{j,k}$. Due to the invariance in eq.(8.13) combined with eq.(8.12), it immediately follows that:

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

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), $M^{\hat{f},A}_{0,k} = M^{X,A}_{0,k} = A * M^X_{0,k} = A * M^{\hat{f}}_{0,k}$. At the odd locations $2k+1$ at the next coarser scale $j = 1$,

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

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

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

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), 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 $(W_\ell)_{\ell=1,\ldots,n}$ are i.i.d., implying that the random variables $(\mathrm{Tr}(\mathrm{Log}(M^W_{j,k}))_{k=0,\ldots,2^j-1}$ 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. 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 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 linear wavelet thresholding

Table 3: 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.

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) 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.