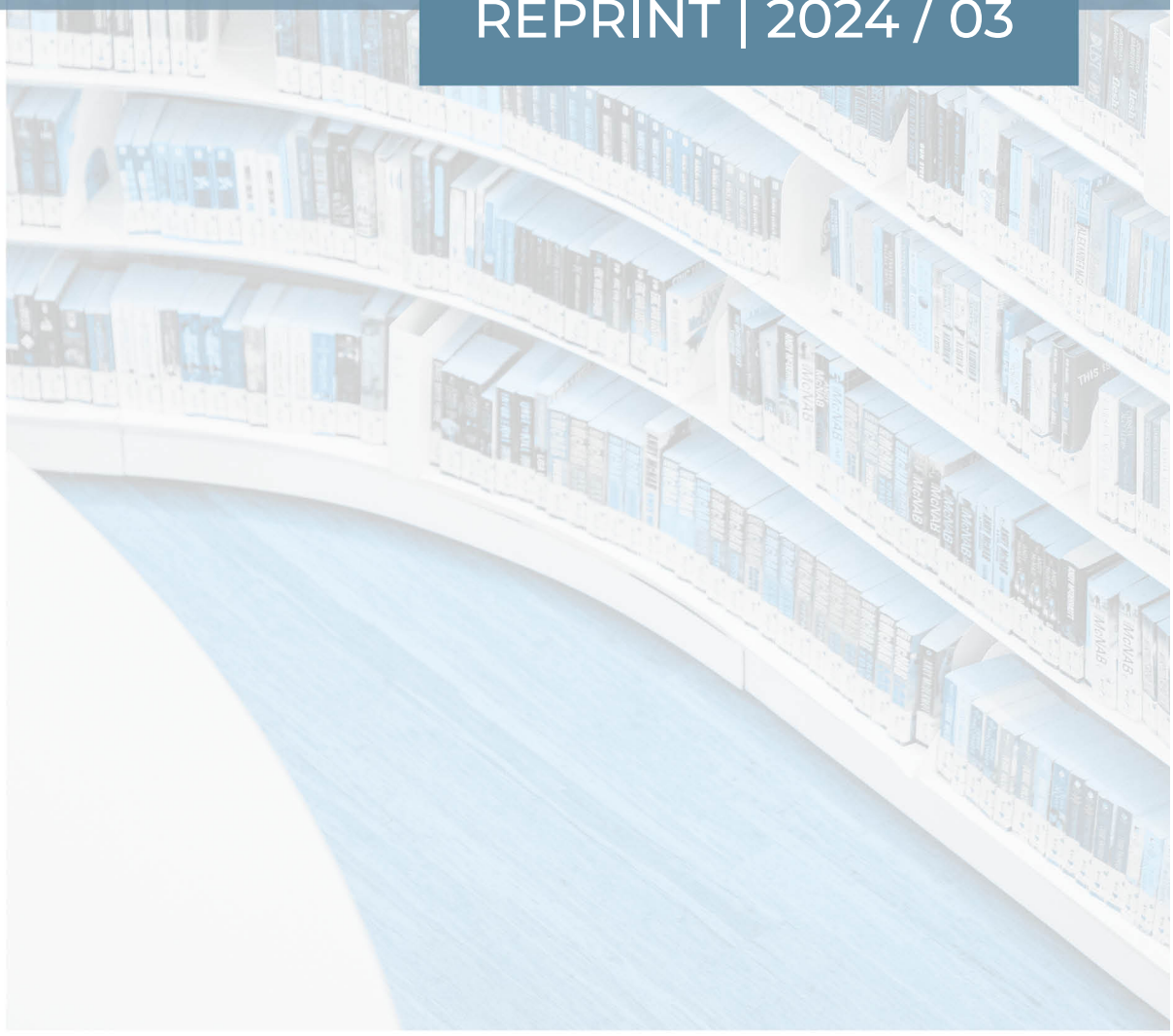
STATISTICAL INFERENCE FOR WAVELET CURVE ESTIMATORS OF SYMMETRIC POSITIVE DEFINITE MATRICES

Daniel Rademacher, Johannes Krebs, Rainer von Sachs

REPRINT | 2024 / 03



UCLouvain

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

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication.html

Statistical inference for wavelet curve estimators of symmetric positive definite matrices

Daniel Rademacher [1] **Johannes Krebs** [2] **Rainer von Sachs** [3]

[1] *Institute of Applied Mathematics, Heidelberg University, 69120 Heidelberg, Germany. E-mail: daniel.rademacher@uni-heidelberg.de*

[2] *Department of Mathematics, KU Eichstätt-Ingolstadt, 85072 Eichstätt, Germany. E-mail: johannes.krebs@ku.de*

[3] *ISBA/LIDAM, UCLouvain, 1348 Louvain-la-Neuve, Belgium. E-mail: rainer.vonsachs@uclouvain.be*

Abstract: In this paper we treat statistical inference for a wavelet estimator of curves of symmetric positive definite (SPD) using the log-Euclidean distance. This estimator preserves positive-definiteness and enjoys permutation-equivariance, which is particularly relevant for covariance matrices. Our second-generation wavelet estimator is based on average-interpolation (AI) and allows the same powerful properties, including fast algorithms, known from nonparametric curve estimation with wavelets in standard Euclidean set-ups. The core of our work is the proposition of confidence sets for our AI wavelet estimator in a non-Euclidean geometry. We derive asymptotic normality of this estimator, including explicit expressions of its asymptotic variance. This opens the door for constructing asymptotic confidence regions which we compare with our proposed bootstrap scheme for inference. Detailed numerical simulations confirm the appropriateness of our suggested inference schemes.

Keywords and phrases: Asymptotic normality, AI refinement, log-Euclidean manifold, SPD matrices, Covariance matrices, Matrix-valued curves, Nonparametric inference, Second generation wavelets.

1. Introduction

In this paper we derive statistical inference for curves of positive definite matrices observed in the presence of noise. Consider a symmetric positive definite (SPD) $d \times d$ matrix where each element is a curve with argument in a domain (e.g. time or space) which is the same for the $d(d+1)/2$ different curves. Prominent examples are covariance matrices that evolve over time (or space), such as, for $d = 3$, diffusion tensors which model the covariance of the Brownian motion of water (or some other liquid) in human tissues (brain, heart, etc) at a voxel. I.e., for each tissue voxel, there is a 3×3 SPD matrix to describe the shape of local diffusion.

Beyond diffusion tensors (Zhu et al. [2009], Dryden et al. [2009]) we find applications, e.g., in computer vision (Caseiro et al. [2012]), medical imaging (Fillard et al. [2007], Fletcher and Joshib [2007]), signal processing (Arnaudson

et al. [2013]) and neuroscience (Friston [2011]). In Chau and von Sachs [2021], the authors considered the slightly larger class of Hermitian positive-definite (HPD) matrices as typically arising in estimation of spectral density matrices of multivariate time series which are functions of frequency. Although HPD matrices are not considered in our work here, it is heavily motivated from Chau and von Sachs [2021] and is using part of their estimation methodology.

As genuinely new contribution of this paper, we address the construction of (pointwise) confidence regions (CR) for curves of SPD matrices (or curve-valued SPD matrices) observed in the presence of noise: we propose both CRs based on asymptotic normality and, as a more practically satisfying alternative, CRs based on a (wild) bootstrap scheme - for which we use the newly proven asymptotic normality for showing its theoretical validity. Note that in order to preserve the PD property we consider our curve-valued matrices as a whole instead of applying inference elementwise on the scalar-valued functions, inference as being based on nonparametric curve estimators (kernels, nearest neighbors, splines, etc.). That means in particular that at each and every point both our curve estimators and our constructed confidence regions must themselves remain in the space of SPD matrices of dimension d which we call $Sym^+(d)$ hereafter.

With our agenda we face the problem that our objects of interest do not live in a vector space which usually gives all the tools - and in particular Euclidean distances - for asymptotic theory of nonparametric curve estimation (such as on rates of convergence of the mean-squared error). To cope with this problem we apply the matrix-logarithm on our SPD curves. This allows us to work in the *vector space* $Sym(d)$ of symmetric $d \times d$ matrices in which we can now derive our statistical estimation and inference. It remains to apply the matrix-exponential to our resulting objects (smoothed curve-valued symmetric matrices), which leads us to preserve positive definiteness (PD) of our resulting curve estimators in $Sym^+(d)$ - without needing to apply any pre- or postprocessing step: Such a device would be prohibitive for a feasible analysis of its statistical properties (in particular mean-square convergence and asymptotic normality of the final estimator).

Our attractively simple construction can, in fact, be seen as an instance of a more general concept of estimation of curves which do not live in a vector space (as in Chau and von Sachs [2021]), the concept of considering *Riemannian manifolds*, i.e. manifolds equipped with a differentiable ("Riemannian") metric (Pennec [2006]): in our particular situation we equip the manifold $Sym^+(d)$ with the *log-Euclidean* metric (introduced by Arsigny et al. [2007]) which yields a simple corresponding Riemannian distance (the Frobenius-norm of the difference of the log SPD-matrices).

Note that there is a growing statistical literature (among others, Yuan et al. [2012] for local polynomial estimation, Rahman et al. [2005], Chau and von Sachs [2021], Chau and von Sachs [2022] for wavelet estimation, and Boumal and Absil [2011], Hinkle et al. [2014] more generally for regression on Rieman-

nian manifolds), that exploits these findings from differential geometry (e.g. Pennec [2006]).

It remains to tell the reader that we choose *wavelet* estimation (see, e.g., Antoniadis et al. [1994]) as our preferred nonparametric curve estimation method, motivated from the following reasons: (i) as known from traditional curve estimation, wavelets are capable to treat curves with inhomogeneous smoothness features (e.g. simultaneous appearance of smooth and locally varying parts such as jumps and cusps); (ii) Chau and von Sachs [2021] delivers a powerful toolbox of how to construct wavelets on non-Euclidean geometries of SPD and HPD matrices via *Average-Interpolation* that we use and recall later.

Our approach on using the log-Euclidean metric is somehow related to other PD-preserving transformations such as taking the Cholesky square root of the considered SPD matrices. Due to the non-uniqueness of the Cholesky square root, this approach is however known to be suboptimal with respect to some desirable invariance properties (Chau and von Sachs [2021], e.g): as the log-Euclidean metric can be shown to be unitary congruence invariant (see also Lemma E.9 in the Supplement Section E), this implies equivariance of our constructed estimators with respect to, e.g., rotations. For estimation of covariance matrices of a multivariate data vector (e.g. a time series) this means in particular that if the coordinates of the entries of the data vector are permuted, the estimator of its covariance matrix follows this permutation - a property lacking for other choices, such as, e.g., the Cholesky and the log-Cholesky metrics (Lin [2019]).

Using the log-Euclidean metric, most importantly, opens a way to derive properly centered asymptotic normality of the proposed curve estimators which is the first of our original contributions here. We will subsequently use this in order to construct correctly centered confidence regions, due to the simplified geometric structure of the log-Euclidean compared to other Riemannian metrics (such as the otherwise very attractive and natural affine-invariant metric). Another advantage is that estimators based on this metric do not suffer from the *swelling effect* - which means that the determinant of an average of two (or more) SPD matrices in this metric is guaranteed to not exceed the determinant of any of the averaged matrices. This property will ensure that our constructed confidence regions which are based on estimators that are constructed as intrinsic averages in $Sym^+(d)$, will not swell towards the borders of this manifold. This is obviously an important advantage for constructing confidence volumes that are not simply based on the Euclidean metric, in order to have them respect the nominal level in practice without becoming incorrectly too large.

The rest of the paper is organized as follows. In the following subsection we give a brief overview of our wavelet estimator and how we achieve statistical inference, omitting all technical details of its construction and properties. In the next section we review some important features of intrinsic *average interpolation* (AI) in the case of the log-Euclidean manifold, as they had been

established in more generality by Chau and von Sachs [2021]. AI schemes, are known to have powerful properties as developed, e.g. in Daubechies and Lagarias [1991, 1992a,b]. Their explicit study allows us to derive precise constants for the asymptotic variance of our newly derived central limit theorem. Moreover Donoho [1993] showed that wavelet estimators based on AI carry the same desirable properties as first-generation wavelet schemes.

The results for wavelet estimation in our nonparametric regression setup of denoising SPD-matrix valued curves are given in Section 3, i.e. reviewing rates of mean-square convergence, and foremost, deriving asymptotic normality including an explicit calculation of the asymptotic variance of our estimators. Section 4 presents the two different constructions of our confidence sets which live in the considered space of SPD-matrix valued curves. Here we base ourselves, on the one hand, on the shown asymptotic normality, and on the other hand, on a newly proposed (wild) bootstrap scheme as the final of our original contributions, for which we show its theoretical validity. Finally we provide numerical simulation studies in Section 5, which show the satisfactory empirical coverage and a comparison between bootstrap confidence regions and those based on asymptotic normality. A short conclusion Section 6 sketches some possible extensions for future work. All proofs, technical details and more background are given in a Supplement.

1.1. Main steps of proposed procedure in a nutshell

To give the reader a first non-technical idea about the achievements of this paper, we describe our procedure of constructing confidence regions for curves of SPD matrices based on Average-Interpolation wavelet schemes briefly in the following.

Recalling classical wavelet smoothing We begin by (very schematically) recalling the first-generation (linear) wavelet estimator in a (demonstratively simple) nonparametric signal-plus-noise regression model, where we observe $n = 2^J$ noisy observations Y_k from a equidistantly sampled square-integrable curve $c(t)$, $t \in [0,1]$, in the presence of (i.i.d.) homoscedastic (mean zero) noise ξ_k:

$$Y_k = c(k/n) + \xi_k, \quad k = 0, \dots, n-1 \ .$$

Projecting each of the ingredients of this additive model on the scaling functions $\{\Phi_{j,k}\}$ of an $L^2([0,1])-$orthonormal wavelet basis furnishes the corresponding additive i.i.d. model in the coefficient domain

$$M_{j,k,n} = M_{j,k} + \xi'_{jk}, \quad k = 0, \dots, 2^j - 1, \ j = 0, \dots, J \ , \tag{1.1}$$

where $M_{j,k,n}$ and $M_{j,k} = \langle c, \Phi_{j,k} \rangle_{L^2}$ denote the empirical and the true scaling coefficients (of the curve c), respectively, on resolution scales $j = 0, \dots, J$, and locations $k = 0, \dots 2^j - 1$. Then a "linearly thresholded" MSE-consistent curve estimator of c is given by

$$\widehat{c}_{J_0}(t) = \sum_k M_{J_0,k,n} \Phi_{J_0,k}(t) \ ,$$

where the smoothing scale J_0 is chosen to be smaller resp. coarser than the sampling scale J, akin the choice of the bandwidth $b \sim 2^{-J_0}$ of the associated (wavelet) kernel estimator to fulfil $b \to 0$ and $nb \to \infty$ with $n \to \infty$.

For the following construction via Average-Interpolation note that although *no* (scaling or wavelet) basis functions $\{\Phi_{j,k}\}$ do exist in the considered non-Euclidean geometry, we will benefit, by equation (3.3), from a signal-to-noise model similar to (1.1) which holds with the same properties as for first-generation wavelet estimation of (scalar-valued) curves. In particular, as it is common practice, the scaling coefficients $M_{J,k,n}$ are identified with the observations Y_k, before a dyadic fast pyramid algorithm (of order $O(n)$) is used to obtain coarser scaling coefficients $M_{j,k,n}$ essentially via successive weighted averaging with filters of finite length.

Construction of the AI wavelet estimator We derive an MSE-consistent linear wavelet estimator now via Average-Interpolation in the space $Sym(d)$, i.e. the estimator is constructed to estimate, in the presence of noise, $\log c(t)$, with $c(t) \in Sym^+(d)$ a (smooth) curve of SPD matrices. For this we essentially follow the construction of Chau and von Sachs [2021], including results on rates of MSE-convergence.
In a nutshell the algorithm follows the same ideas from the aforementioned dyadic pyramid scheme. With this we obtain the empirical scaling coefficients, now $\log M_{J,k,n}$ on the sample scale J from observed sampled values of the curve $\log c(k/2^J), k = 0, \dots 2^J - 1$, corrupted by noise. A simple but not the only possible signal-plus-noise model is similar to equation (1.1), now of the form $\log M_{j,k,n} = \log M_{j,k} + \xi''_{jk}$, with ξ''_{jk} following a noise distribution in $Sym(d)$ (e.g. a log-normal distribution, see Example 3 in Section 4). This construction allows to apply the same paradigms of wavelet denoising as in the traditional scalar case. Linearly thresholding all empirical wavelet coefficients $\log M_{j,k,n}$ for scales $j > J_0$ is again essentially equivalent to a kernel estimator of bandwidth 2^{-J_0} (see, e.g., Antoniadis et al. [1994]). Asymptotically, considering the common nonparametric smoothing conditions $J_0 \to \infty$ with $J \to \infty$ while $J - J_0 \to \infty$, delivers an MSE-consistent estimator with the usual optimal rate of convergence, depending on the smoothness regularity of the underlying target curve $c(t)$ (measured in the space of SPD matrices).

Confidence regions for SPD-matrix valued curves based on asymptotic normality Using the construction of the AI wavelet estimator, we propose a novel approach for the construction of confidence regions for $c(t)$ which are based on asymptotic normality - which we show to hold for the collection of the empirical scaling coefficients $\log M_{j,k,n}$ with the usual nonparametric rate of variance decaying to zero as $2^{-(J-j)}$. We derive the correct asymptotic variance-covariance structure of these CLTs via the study of an appropriate covariance operator for matrix-valued objects. Through a connection between the construction of AI-schemes and the study of convergence of these schemes in Daubechies and Lagarias [1991, 1992a,b] we establish the limiting behaviour (for $J_0 \to \infty$

as $n = 2^J \to \infty$) of the variance of our smoothed wavelet estimator $\log M_{J_0,k,n}$.

Confidence regions based on bootstrap As second new ingredient we propose confidence regions based on a wild bootstrap approach where we simulate bootstrap residuals from our wavelet estimator considered above. Using an appropriate signal-plus-noise model in $Sym(d)$, we thus create bootstrap replicates of our original observations on the finest scale J and apply the same wavelet estimation schemes as in the original world. This bootstrap scheme is shown to be asymptotically valid, and turns out to give the same satisfactory empirical results in our extensive simulation studies as the regions based on asymptotic normality - with both their empirical coverages being close to keeping the nominal level.

2. The intrinsic AI wavelet transform

This section is largely inspired by the developments in Chau and von Sachs [2021], adapted to the particular situation of a log-Euclidean manifold.

We start out by briefly discussing some fundamentals as well as introducing some necessary notation. Further details on the geometric aspects of the present setting, including the Fréchet mean for the log-Euclidean metric, are provided in the Supplement, Section A.1. In the second part we then present the *average-interpolation* (AI) approach as introduced in the aforementioned paper in the context of a log-Euclidean manifold (for classical wavelet AI we refer to Klees and Haagmans [2000] and to Donoho [1993]). Generally, AI for wavelets provide multiscale representations and can successfully be implemented via *refinement schemes*. We may emphasize that an important aspect of the AI approach of Chau and von Sachs [2021] is that, in order to implement the *predicting step* that is inherent to all second generation schemes based on *lifting* (Jansen and Oonincx [2005]), predicting the midpoints of the refinement scheme relies on *intrinsic average* interpolation. In the third part, we detail the concept of forward and backward average interpolation (which provide the fast wavelet transform), hence the multiscale algorithm for processing observed data and their estimated local averages across different scales. Finally, in the fouth part, we complement the findings of Chau and von Sachs [2021] and discuss convergence of AI refinement schemes.

2.1. Preliminaries

We begin by collecting some notation: Let $M(d)$ denote the space of real $d \times d$ matrices and $GL(d)$ the subspace of invertible matrices. Then for any $A \in M(d)$, its *matrix exponential* is given by $\exp(A) = \sum_{k=0}^{\infty} A^k/k! \in GL(d)$. In particular the restriction $\exp\colon Sym(d) \to Sym^+(d)$ is one-to-one and for any $S \in Sym^+(d)$, there is a $\log(S) = A \in Sym(d)$ such that $\exp(A) = S$. The mapping $\log\colon Sym^+(d) \to Sym(d)$ is called *(principal) matrix logarithm*. We write $\langle A, B\rangle_F := \mathrm{trace}(AB)$, for the Frobenius inner product of two matrices

$A, B \in M(d)$ and $\|\cdot\|_F$ for the induced norm. We may also omit the sub index depending on the context.

As outlined in the introduction, this work is concerned with intrinsic curve estimation within the manifold $Sym^+(d)$ of $d \times d$ symmetric positive definite (SPD) matrices. Now there are several ways to equip this differentiable manifold with a Riemannian metric, that is a positive definite inner product g_S on the tangent space $T_S Sym^+(d)$ defined at each point $S \in Sym^+(d)$ and depending smoothly on S. We remark that $T_S Sym^+(d)$ can be identified with the ambient vector space $Sym(d)$ of symmetric matrices in the sense that both spaces are isomorphic. To circumvent some delicate issues that arise in inferential investigations on curved spaces, we choose to consider the log-Euclidean metric, which was introduced in Arsigny et al. [2007] and provides a *flat geometry* on $Sym^+(d)$. A compact exposition of the Riemannian structure induced by the log-Euclidean metric is provided in the Supplement, Section E. Although the metric itself has a rather complicated expression, the induced distance is simply given by

$$d(S_1, S_2) = \|\log(S_2) - \log(S_1)\|_F, \quad S_1, S_2 \in Sym^+(d), \tag{2.1}$$

which turns out to be quite advantageous in view of statistical derivations.

Turning to the probabilistic aspects, we recall, from the Hopf-Rinow theorem (cf. Pennec [2006]), that $(Sym^+(d), d)$ is a complete separable metric space. Let $\mathcal{B}(Sym^+(d))$ denote the corresponding (w.r.t. the distance d) Borel σ-algebra, then we call a measurable function $X\colon (\Omega, \mathcal{A}, \mathbb{P}) \to (Sym^+(d), \mathcal{B}(Sym^+(d)))$ a random element on $Sym^+(d)$ and the induced probability measure $\mathbb{P}^X := \mathbb{P} \circ X^{-1}$ on $Sym^+(d)$ is as usual referred to as the law or distribution of X. Furthermore, let $\mathcal{P}(Sym^+(d))$ denote the set of all probability measures on $(Sym^+(d), \mathcal{B}(Sym^+(d)))$ and $\mathcal{P}_r(Sym^+(d))$ the subset of probability measures having a finite moment of order r with respect to the log-Euclidean distance (2.1), i.e.

$$\mathcal{P}_r(Sym^+(d)) := \left\{ \nu \in \mathcal{P}(Sym^+(d))\colon \int_{Sym^+(d)} d(S_0, S)^r \,\mathrm{d}\nu(S) < \infty \right\}, \tag{2.2}$$

where $S_0 \in Sym^+(d)$ is arbitrary. We note in passing that we will use λ to denote the Lebesgue measure on $\mathbb{R}$.

2.2. Intrinsic average-interpolation refinement scheme

The data are assumed to stem from a curve $c\colon [0,1] \to Sym^+(d)$, which is square integrable in that $\int_0^1 \|c(t)\|_F^2 \,\mathrm{d}t < \infty$. Following the usual paradigm in nonparametric wavelet regression, the $n = 2^J$ data observations $c_k = c(k2^{-J})$, $k = 0, \ldots, 2^J - 1$ result from (noisy versions of) equidistant sampling of the curve $c(t)$ on a (dyadic) resolution scale J, i.e. for grid points $t_k = k/2^J$. As a common choice for refinement schemes based on wavelets, those sampled data points are identified with the *scaling coefficients* or *midpoints* $(M_{J,k})_k$ of the

(wavelet) refinement scheme on the sampling scale J. That is, the *input data* of the refinement scheme are the $(M_{J,k})_k$, observed on scale J, and they are assumed to represent the average of the curve (see (A.3) in the Supplement) over a dyadic interval

$$M_{J,k} = \mathrm{Ave}_{2^{-J}[k,k+1)}(c) = \exp\left(2^J \int_{k2^{-J}}^{(k+1)2^{-J}} \log(c(t))\,\mathrm{d}t\right), \quad k = 0, \dots, 2^J - 1,$$

where $2^{-J}[k, k+1) := [k/2^J, (k+1)/2^J)$ forms a uniform partition of $[0, 1)$. We note that the intrinsic refinement scheme described in the following can also be adapted to a *non-dyadic* observation grid, see for instance Chau [2018].

In order to later utilize the refinement scheme for regression analysis the requirements for the scheme are twofold. On the one hand, refined or predicted midpoints have to form an average of coarser scale midpoints, so the transformation is suitable for noise-removal. On the other hand, the success of the refinement schemes hinges on the ability to reconstruct what is to be defined an *intrinsic polynomial* of a given degree without any error of approximation, akin the essential property that is enjoyed by wavelet schemes (of adapted order) in the classical setting. For a formal definition of intrinsic polynomials we refer to Hinkle et al. [2014]. As already mentioned, Chau and von Sachs [2021] derived a suitable refinement scheme that fits those requirements by adapting the classical average-interpolation (AI) refinement scheme as introduced in Donoho [1993]. The basic idea is that based on j-scale midpoints $(M_{j,k})_k$ the predicted or refined $(j+1)$-scale midpoints, denoted $(\widetilde{M}_{j+1,k})_k$, are computed as the $(j+1)$-scale midpoints of unique intrinsic polynomials with j-scale midpoints $(M_{j,k})_k$. To elaborate, let $N = 2L+1$ for some $L \geq 0$ and fix a location $k \in \{L, \dots, 2^j - L\}$. The *intrinsic AI-refinement scheme of order or degree* N is then formally defined as follows:

Let $\rho = \rho_{j,k} \colon [0,1] \to Sym^+(d)$ be the unique intrinsic polynomial of degree $N-1$ with j-scale midpoints $M_{j,k-L}, \dots, M_{j,k+L}$, that is

$$\mathrm{Ave}_{2^{-j}[k',k'+1)}(\rho) = \exp\left(2^j \int_{k'2^{-j}}^{(k'+1)2^{-j}} \log(\rho(t))\,\mathrm{d}t\right) = M_{j,k'}, \quad k' = k-L, \dots, k+L.$$

Then calculate the two predicted midpoints $\widetilde{M}_{j+1,2k}$ and $\widetilde{M}_{j+1,2k+1}$ on the next finer scale as

$$\widetilde{M}_{j+1,2k} = \mathrm{Ave}_{2^{-(j+1)}[2k,2k+1)}(\rho) = \exp\left(2^{j+1} \int_{k2^{-j}}^{(2k+1)2^{-(j+1)}} \log(\rho(t))\,\mathrm{d}t\right), \tag{2.3}$$

$$\widetilde{M}_{j+1,2k+1} = \mathrm{Ave}_{2^{-(j+1)}[2k+1,2k+2)}(\rho) = \exp\left(2^{j+1} \int_{(2k+1)2^{-(j+1)}}^{(k+1)2^{-j}} \log(\rho(t))\,\mathrm{d}t\right). \tag{2.4}$$

Computing the integrals on the right hand side directly is of course infeasible, but similarly to AI-refinement on the real line, the AI refinement scheme (2.3) and (2.4) reduces in practice to some elementary linear algebra (see Chau and von Sachs [2021]), which can be encoded in a vector of *prediction coefficients* $(c_1, \ldots, c_L)$, i.e.

$$\begin{aligned} \widetilde{M}_{j+1,2k} &= \mathrm{Ave}\left(\{M_{j,k-L}, \ldots, M_{j,k+L}\},\ \{-c_L, \ldots, -c_1, 1, c_1, \ldots, c_L\}\right), \\ \widetilde{M}_{j+1,2k+1} &= \mathrm{Ave}\left(\{M_{j,k-L}, \ldots, M_{j,k+L}\},\ \{c_L, \ldots, c_1, 1, -c_1, \ldots, -c_L\}\right). \end{aligned} \tag{2.5}$$

Here we make use of the abbreviation (cf. (A.2) in the Supplement)

$$\mathrm{Ave}(\{S_i\}; \{w_i\}) := \exp\left(\sum_{i=1}^{n} w_i \log(S_i)\right), \qquad S_i \in Sym^+(d).$$

In fact, these prediction coefficients are the same as in the average-interpolation transform in the Euclidean case (see Donoho [1993]), since essentially only classical interpolation is replaced with intrinsic interpolation. For example, if $L = 1$ then $c_1 = -1/8$, if $L = 2$ then $(c_1, c_2) = (-22, 3)/128$ and if $L = 3$ then $(c_1, c_2, c_3) = (-201, 44, -5)/1024$.

2.3. Intrinsic forward and backward average-interpolation wavelet transform

The intrinsic AI-refinement scheme developed in the previous section naturally leads to a corresponding *intrinsic average-interpolation wavelet transform* that can be used to synthesize a smooth curve in $Sym^+(d)$. The first step in constructing such a wavelet transform is to build up a redundant midpoint (or scaling coefficient) pyramid. The midpoint pyramid algorithm aggregates information from the finest observation scale J successively to coarser scales. It is at the heart of the wavelet transform.

Midpoint pyramid. Consider scaling coefficients $M_{J,0}, \ldots, M_{J,2^J-1}$ at the *finest scale* J to be given by the sampled data. At the next coarser scale $j = J-1$ set $M_{j,k}$ to be the halfway point or *midpoint* on the geodesic γ (see (E.4) in the Supplement) passing through $M_{j+1,2k}$ and $M_{j+1,2k+1}$, i.e.

$$M_{j,k} := \gamma(1/2; M_{j+1,2k}, M_{j+1,2k+1}) = \mathrm{Ave}(M_{j+1,2k}, M_{j+1,2k+1}) \tag{2.6}$$

for $k = 0, \ldots, 2^j - 1$. Continue this coarsening operation up to scale $j = 0$ to obtain the *midpoint pyramid* $((M_{j,k})_{j,k} : 0 \le j \le J,\ 0 \le k \le 2^j - 1)$, see also the scheme in Figure 6 in the Supplement.

Now the above considerations lead to the following forward ($j \rightsquigarrow j-1$) and backward ($j-1 \rightsquigarrow j$) AI-wavelet transform:

Forward wavelet transform. As *input* take the j-scale midpoints $(M_{j,k})_k$. The *output* are the $(j-1)$-scale midpoints $(M_{j-1,k})_k$ and j-scale *wavelet coefficients* $(D_{j,k})_k$. In detail, the forward transform is performed as follows:

1. *Coarsen.* (i) Compute the $(j-1)$-scale midpoints $(M_{j-1,k})_k$ according to the midpoint relation (2.6).
   (ii) Select a refinement order $N = 2L+1$, $L \geq 0$, and generate predicted midpoints $(\widetilde{M}_{j,k})_k$ based on $(M_{j-1,k})_k$ via (2.5).
2. *Difference.* Define the *wavelet coefficients* $(D_{j,k})_k$ as an intrinsic difference according to

$$D_{j,k} := 2^{-j/2}\mathrm{Log}_{\widetilde{M}_{j,2k+1}}(M_{j,2k+1}) \in T_{\widetilde{M}_{j,2k+1}}Sym^+(d). \tag{2.7}$$

Backward wavelet transform. As *input* take the $(j-1)$-scale midpoints $(M_{j-1,k})_k$ and j-scale wavelet coefficients $(D_{j,k})_k$. The *output* are the j-scale midpoints $(M_{j,k})_k$. In detail, the backward transform is performed as follows:

1. *Refine.* (i) Given the refinement order $N = 2L+1$, $L \geq 0$, generate odd predicted midpoints $(\widetilde{M}_{j,2k+1})_k$ based on $(M_{j-1,k})_k$ via (2.5).
   (ii) Compute the odd j-scale midpoints $(M_{j,2k+1})_k$ by reversing the relation (2.7), i.e.,

$$M_{j,2k+1} = \mathrm{Exp}_{\widetilde{M}_{j,2k+1}}(2^{j/2}D_{j,k}). \tag{2.8}$$

2. *Complete.* Compute the even j-scale midpoints $(M_{j,2k})_k$ through the midpoint relation (2.6), i.e.,

$$M_{j,2k} = \exp\left(2\log(M_{j-1,k}) - \log(M_{j,2k+1})\right). \tag{2.9}$$

Thus, given the coarsest midpoint $M_{0,0}$ and the wavelet coefficient pyramid $(D_{j,k})_{j,k}$ for $j = 0,\dots,J$ and $k = 0,\dots,2^{j-1}-1$, the original input sequence $(M_{J,k})_k$ can be retrieved by consecutively repeating the reconstruction procedure (2.8) and (2.9) up to scale J.

Remark 2.1. By construction, the wavelet coefficients $D_{j,k}$ live in different tangent spaces. In order to compare these quantities (e.g., with a given uniform threshold), parallel transporting them to the tangent space at the identity results in so called *whitened wavelet coefficients* given by

$$\begin{aligned}\mathfrak{D}_{j,k} &:= \Gamma^{\mathrm{Id}}_{\widetilde{M}_{j,2k+1}}(D_{j,k}) \\ &= 2^{-j/2}\big(\log(M_{j,2k+1}) - \log(\widetilde{M}_{j,2k+1})\big) \in T_{\mathrm{Id}}Sym^+(d) = Sym(d).\end{aligned} \tag{2.10}$$

Of course the "length" of the vectors does not change, that is

$$g_{\widetilde{M}_{j,2k+1}}(D_{j,k}, D_{j,k}) = 2^{-j}\left\|\log(M_{j,2k+1}) - \log(\widetilde{M}_{j,2k+1})\right\|_F^2 = \|\mathfrak{D}_{j,k}\|_F^2.$$

2.4. Convergence of AI refinement schemes

The estimation procedure which we introduce in the subsequent section 3 will require multiple consecutive applications of the AI refinement scheme when lifting the wavelet estimator from a coarse scale J_0 to a finer scale J. To manage such iterative transformations it turns out to be convenient to express the averaging operations (2.5) via suitable transition matrices. This will in particular allows us to study the asymptotic behaviour of our wavelet estimator. To elaborate, let us first define for a generic $m \times n$ matrix $A = (a_{ij})_{i,j}$, the blocked $md \times nd$ matrix $A_{\diagdown d}$ by

$$A_{\diagdown d} = (a_{ij} I_d)_{i,j},$$

where $I_d \in \mathbb{R}^{d\times d}$ is the identity matrix. This means $A_{\diagdown d}$ consists of block matrices each of these matrices being the $a_{i,j}$ multiple of the identity I_d. Next, for a refinement order $N = 2L+1$, we can use the prediction coefficients $(c_1, \dots, c_L)$ to define the matrices $E_N, O_N \in \mathbb{R}^{(2N-1)\times(2N-1)}$, according to ($E_N$) and ($O_N$) at the end of Section A of the Supplement. Notice that E_N and O_N are similar as $O_N = ZE_NZ$, where $Z = [e_{2N-1}, \dots, e_1]$ and e_i being the ith standard basis (column) vector in $\mathbb{R}^{2N-1}$. A careful observation then reveals that the AI-refinement (2.5) can also be described with $(E_N)_{\diagdown d}$ and $(O_N)_{\diagdown d}$ via

$$\begin{bmatrix} \log \widetilde{M}_{j+1,2k-2L} \\ \vdots \\ \log \widetilde{M}_{j+1,2k+2L} \end{bmatrix} = (E_N)_{\diagdown d} \begin{bmatrix} \log M_{j,k-2L} \\ \vdots \\ \log M_{j,k+2L} \end{bmatrix}, \tag{2.11}$$

respectively

$$\begin{bmatrix} \log \widetilde{M}_{j+1,2k-2L+1} \\ \vdots \\ \log \widetilde{M}_{j+1,2k+2L+1} \end{bmatrix} = (O_N)_{\diagdown d} \begin{bmatrix} \log M_{j,k-2L} \\ \vdots \\ \log M_{j,k+2L} \end{bmatrix}. \tag{2.12}$$

Consequently, by a successive application of (2.11) and (2.12), we derive the following:

Lemma 2.2. *Let $m, j \geq 1$ and $p \in \{0, \dots, 2^m - 1\}$ with unique binary representation $p = \sum_{i=j+1}^{j+m} d_i(p) 2^{j+m-i}$, where $d_i(p) \in \{0, 1\}$. Then*

$$\begin{bmatrix} \log \widetilde{M}_{j+m,2^m k+p-2L} \\ \vdots \\ \log \widetilde{M}_{j+m,2^m k+p+2L} \end{bmatrix} = (U_m U_{m-1} \cdot \ldots \cdot U_1)_{\diagdown d} \begin{bmatrix} \log M_{j,k-2L} \\ \vdots \\ \log M_{j,k+2L} \end{bmatrix}$$

with

$$U_i = E_N \mathbb{1}\{d_{j+i}(p) = 0\} + O_N \mathbb{1}\{d_{j+i}(p) = 1\}, \quad i = 1, \dots, m.$$

So, it is of crucial interest whether arbitrary products $U_m U_{m-1} \dots U_1$ of transition matrices $U_i \in \{E_N, O_N\}$ converge as m increases to infinity. The answer is affirmative and follows from results in Daubechies and Lagarias [1991, 1992b] as well as Donoho [1993]. To see this, let us first define

$$(\widehat{c}_{-2L}, \dots, \widehat{c}_{2L+1}) := (c_L, -c_L, \dots, c_1, -c_1, 1, 1, -c_1, c_1, \dots, -c_L, c_L) \ .$$

The unique non-trivial solution of the two-scale difference equation

$$f(x) = \sum_k \widehat{c}_k \, f(2x-k), \qquad f \in L^1 \tag{2.13}$$

is then given by the limit (that is an application ad infinitum) of the classical AI refinement scheme on the real line to the Kronecker sequence $(\delta_{0,k})_k$, see Donoho [1993] Theorem 2.1 and Daubechies and Lagarias [1992b] Theorem 2.2 for a detailed explanation. Let φ_L denote this *fundamental solution* of (2.13). For further properties of φ_L, in particular its regularity, we refer again to the aforementioned papers. Now the connection between φ_L and the limit of products of the form $U_m U_{m-1} \cdots U_1$, is as follows: For any $x \in [0,1]$ consider its binary expansion

$$x = \sum_{j=1}^{\infty} d_j(x) 2^{-j}, \qquad d_j(x) \in \{0,1\}.$$

Adapting the notation in Daubechies and Lagarias [1992b] let us define

$$T_0 = Z E_N^T Z \quad \text{and} \quad T_1 = Z O_N^T Z. \tag{2.14}$$

Then as a consequence of Theorem 2.2 in Daubechies and Lagarias [1992b] we have that all infinite products of T_0 and T_1 converge to the limit

$$\lim_{m\to\infty} T_{d_1(x)} T_{d_2(x)} \cdot \ldots \cdot T_{d_m(x)} = \begin{bmatrix} \varphi_L(x-2L) & \cdots & \varphi_L(x-2L) \\ \vdots & & \vdots \\ \varphi_L(x+2L) & \cdots & \varphi_L(x+2L) \end{bmatrix},$$

where $(d_j(x))_j$ are just the coefficients in the binary expansion of $x \in [0,1]$. Consequently, if we let

$$U_{d_j(x)} = E_N \, \mathbb{1}\{d_j(x) = 0\} + O_N \, \mathbb{1}\{d_j(x) = 1\},$$

then due to (2.14) the above convergence immediately implies, for each $x \in [0,1]$,

$$\lim_{m\to\infty} \left(U_{d_m(x)} U_{d_{m-1}(x)} \cdot \ldots \cdot U_{d_1(x)}\right)_{\diagdown d} = \begin{bmatrix} \varphi_L(x-2L) & \cdots & \varphi_L(x+2L) \\ \vdots & & \vdots \\ \varphi_L(x-2L) & \cdots & \varphi_L(x+2L) \end{bmatrix}_{\diagdown d} =: \Phi_L(x)_{\diagdown d}. \tag{2.15}$$

3. Wavelet regression for smooth curves in $Sym^+(d)$

Having introduced in the previous section the AI-wavelet scheme for SPD-matrix valued data, we can now describe our procedure of how to estimate smooth curves of SPD matrices, observed in the presence of noise. For this, let us first collect and, in a sufficient manner, complement the conditions given so far, so that the setup we are operating in is fixed.

Assumption 3.1 (r). Let $c\colon [0,1] \to Sym^+(d)$ be a square integrable function in the sense that $\int_0^1 \|c(t)\|_F^2 \, dt < \infty$. For $J \geq 0$ and $n = 2^J$ set $c_k := c(k2^{-J})$ as well as

$$M_{J,k} = \mathrm{Ave}_{2^{-J}[k,k+1)}(c) = \exp\left(2^J \int_{k2^{-J}}^{(k+1)2^{-J}} \log(c(t)) \, \mathrm{d}t\right), \quad k = 0, \ldots, n-1.$$

As $(c_k)_k$ respectively $(M_{J,k})_k$ are only observed in the presence of noise, we study an independent sample of scaling coefficients $M_{J,0,n}, \ldots, M_{J,n-1,n} \in Sym^+(d)$ such that, for some $r \geq 1$,

(i) $\mathbb{E}[M_{J,k,n}] = M_{J,k}$ and $\nu_{J,k} := \mathbb{P}^{M_{J,k,n}} \in \mathcal{P}_r(Sym^+(d))$
(ii) the logarithmized scaling coefficients are uniformly bounded in the sense that for some constant $K(r) < \infty$,

$$\sup_{\substack{k=0,\ldots,2^J-1,\\ J\geq 0}} \mathbb{E}\,\|\log M_{J,k,n}\|_F^r \leq K(r). \tag{3.1}$$

Moreover, to be consistent, we assume that the distributions of $M_{J,k,n}$ and $M_{J',k',n'}$ are equal whenever $k2^{-J}$ and $k'2^{-J'}$ represent the same dyadic number, that is,

(iii) the family of corresponding probability measures $\{\nu_{J,k} \mid k = 0, \ldots, 2^J - 1 \text{ and } J \geq 0\}$, satisfies

$$\nu_{J,k} = \nu_{J',k'}, \text{ whenever } k2^{-J} = k'2^{-J'}. \tag{3.2}$$

As an example consider the following intrinsic Riemannian signal plus noise model:

Example 1 (Riemannian signal plus noise model)**.** *Suppose $c\colon [0,1] \to Sym^+(d)$ is an unknown target signal such that $\sup_t \|c(t)\|_F < \infty$ and set $c_k = c(k/n)$ for $k = 0, \ldots, n-1$, $n = 2^J$. Due to the lack of a linear structure, we can not, contrary to the Euclidean case, model a disturbance of the signal by directly adding an independent noise at sample points c_k. Instead, one possible approach is to generate i.i.d. noise vectors $(\xi_k)_k$ in the tangent space $T_{\mathrm{Id}}(Sym^+(d)) \cong Sym(d)$, then parallel transport them to the tangent space $T_{c_k}(Sym^+(d))$ and finally apply the exponential mapping $\mathrm{Exp}_{c_k}\colon T_{c_k}(Sym^+(d)) \to Sym^+(d)$ to model a stochastic disturbance at c_k. Taking into account the closed form expressions of these*

operations (see table 7 in the Supplement), this leads to the following model: For $k = 0, \dots, n-1$, *let*

$$M_{J,k,n} = \mathrm{Exp}_{c_k}\left(\Gamma_{\mathrm{Id}}^{c_k}(\xi_k)\right) = \exp\big(\log(c_k) + \xi_k\big), \tag{3.3}$$

where $(\xi_k)_k$ *is an i.i.d. mean-zero noise in* $Sym(d)$ *such that* $\mathbb{E}\|\xi_k\|_F^r < \infty$ *for some* $r \geq 1$. *It follows immediately (see (A.1) in the Supplement) that*

$$\mathbb{E}[M_{J,k,n}] = \exp\left(\mathbb{E}[\log M_{J,k,n}]\right) = \exp(\log(c_k) + \mathbb{E}[\xi_k]) = c_k = M_{J,k},$$

where the last equality is due to the common habit in wavelet curve estimation, to identify the sampled curve observations c_k *with the scaling coefficients* $M_{J,k}$. *Moreover, we have*

$$\sup_{J,k}\ \mathbb{E}\|\log M_{J,k,n}\|_F^r \leq 2^{r-1}\left(\sup_t \|\log c(t)\|_F^r + \mathbb{E}\|\xi_0\|_F^r\right) =: K(r) < \infty.$$

Let $(\nu_{J,k})_{J,k}$ *denote the family of distributions corresponding to* $(M_{J,k,n})_{J,k}$ *for* $k = 0, \dots, 2^J - 1$ *and* $J \geq 0$. *Then, for an arbitrary* $S_0 \in Sym^+(d)$,

$$\begin{aligned}
\int_{Sym^+(d)} d(S_0, S)^r \,\mathrm{d}\nu_{J,k}(S) &\leq 2^{r-1}\|\log(S_0)\|_F^r + 2^{r-1}\int_{Sym^+(d)} \|\log(S)\|_F^r \,\mathrm{d}\nu_{J,k}(S) \\
&= 2^{r-1}\|\log(S_0)\|_F^r + 2^{r-1}\int_{Sym(d)} \|S\|_F^r \,\mathrm{d}\mathbb{P}^{\log(M_{J,k,n})}(S) \\
&= 2^{r-1}\|\log(S_0)\|_F^r + 2^{r-1}\int_{Sym(d)} \|S + \log(c_k)\|_F^r \,\mathrm{d}\mathbb{P}^{\xi_k}(S) \\
&\leq 2^{r-1}\|\log(S_0)\|_F^r + 2^{2r-2}\left(\|\log(c_k)\|_F^r + \mathbb{E}\|\xi_k\|_F^r\right),
\end{aligned}$$

which shows $\nu_{J,k} \in \mathcal{P}_r(Sym^+(d))$. *Also, (3.3) clearly implies* $\nu_{J,k} = \nu_{J',k'}$, *whenever* $k2^{-J} = k'2^{-J'}$. *We can therefore conclude, that Assumption 3.1 (r) is satisfied for the intrinsic Riemannian signal plus noise model.*

Next, we present an algorithmic description of the estimation procedure, based on the AI forward and backward wavelet transform presented in Section 2.3:

The estimation procedure. Let $X = (X_k)_{k=0,\dots,n-1}$ be a sequence of SPD matrices sampled on a dyadic scale J such that $2^J = n$.

Step 1. Set $M_{J,k,n} = X_k$ for $k = 0, \dots, n-1$ and let $(M_{j,k,n})_{j,k}$ denote the *midpoint pyramid* according to (2.6).

Step 1. Choose a refinement order $N = 2L + 1$, $L \geq 0$, and apply the *forward wavelet transform* to obtain predicted midpoints $(\widetilde{M}_{j,k,n})_{j,k}$ as well as empirical wavelet coefficients $(\widehat{D}_{j,k,n})_{j,k}$.

Step 3. Choose a *thresholding scale* $J_0 < J$ and define thresholded empirical wavelet coefficients $\widehat{D}^{(t)}_{j,k,n}$ according to

$$\widehat{D}^{(t)}_{j,k,n}(J_0) := \begin{cases} \widehat{D}_{j,k,n} & , \ j = 0, \dots, J_0 \\ g(\widehat{D}_{j,k,n}) & , \ j = J_0 + 1, \dots, J \end{cases} \quad \text{for } k = 0, \dots, 2^j - 1\ , \tag{3.4}$$

where $g\colon Sym(d) \to Sym(d)$ is an appropriate thresholding function (shrinking to zero, or setting to zero its argument, as a function of a threshold to be determined from the data).

Step 4. Given the empirical midpoint pyramid $(M_{j,k,n})_{j,k}$ and the thresholded empirical wavelet coefficients $(\widehat{D}^{(t)}_{j,k,n})_{j,k}$, recursively apply the *backward wavelet transform* to obtain estimated midpoints $(\widehat{M}_{j,k,n}(J_0))_{j,k}$. At the sampling scale J, utilize the *wavelet estimates* $(\widehat{M}_{J,k,n}(J_0))_k$ to construct the (curve-)estimator

$$\hat{c}_n(t) = \sum_{k=0}^{n-1} \widehat{M}_{J,k,n}(J_0)\mathbb{1}_{2^{-J}[k,k+1)}(t), \quad t \in [0,1]. \tag{3.5}$$

Remark 3.2 (Linear wavelet estimator). Note that by choosing $g(\cdot) \equiv 0$ in (3.4), we include *linear* thresholding, which amounts to a simple linear smoother, constructed from the midpoints $(M_{J_0-1,k,n})_k$ and lifted back on the sample scale J by successive averaging via (2.5): In more detail, the *linear wavelet estimator* $(\widehat{M}_{J,k,n}(J_0))_k$ is obtained recursively, for $j = J_0+1, \dots, J$ and $k = 0, \dots, 2^j - 1$, via

$$\begin{aligned}
\widehat{M}_{j,2k+1,n}(J_0) &= \mathrm{Ave}\left(\{\widehat{M}_{j-1,k-L,n}(J_0), \dots, \widehat{M}_{j-1,k+L,n}(J_0)\}, \{c_L, \dots, c_1, 1, -c_1, \dots, -c_L\}\right), \\
\widehat{M}_{j,2k,n}(J_0) &= \exp\left(2\log \widehat{M}_{j-1,k,n} - \log \widehat{M}_{j,2k+1,n}\right) \\
&= \mathrm{Ave}\left(\{\widehat{M}_{j-1,k-L,n}(J_0), \dots, \widehat{M}_{j-1,k+L,n}(J_0)\}, \{-c_L, \dots, -c_1, 1, c_1, \dots, c_L\}\right),
\end{aligned} \tag{3.6}$$

with $\widehat{M}_{J_0,k,n}(J_0) = M_{J_0,k,n}$. In the sequel of this work, our theoretical investigations will concentrate on this specific linear estimator.

Remark 3.3 (Non-linear thresholding). A common *non-linear* thresholding scheme would be to choose a keep-or-kill rule of the form

$$g(\widehat{D}_{j,k,n}) = \widehat{D}_{j,k,n}\mathbb{1}\left\{h(\widehat{\mathfrak{D}}_{j,k,n}) > \varepsilon\right\}$$

for some $\varepsilon > 0$ and an appropriate functional $h : Sym(d) \to [0, \infty)$. A convenient choice (see Chau and von Sachs [2021]) is for example $h(\cdot) \equiv |\mathrm{trace}(\cdot)|$.

For the remainder of this section, we now concentrate on establishing limit properties of the linear wavelet estimator (3.6). Of course, as random object within the manifold $Sym^+(d)$ of SPD-matrices, limiting distributional properties of $\widehat{M}_{J,k,n}$ are hard to investigate directly. To circumvent difficulties arising from the non-linearity of the space and the lack of a Gaussian distribution in the classical sense, we study instead the logarithmized coefficients $\log \widehat{M}_{J,k,n}(J_0)$, for which the classical machinery of asymptotic statistics can then be applied again.

The covariance operator. In order to derive asymptotic normality and confidence regions for the logarithmized wavelet estimates $\log \widehat{M}_{J,k,n}(J_0)$, we

need to first model the covariance structure of the sampling scheme. To that end, suppose Assumption 3.1(r) holds for some $r \geq 2$. Then, by the uniform bounded moments condition (3.1) the covariance operator of $\log M_{J,k,n}$ exists and is given by

$$\mathcal{C}(k2^{-J}) := \mathrm{Cov}(\log M_{J,k,n}) \colon \begin{cases} Sym(d) & \to Sym(d), \\ A & \mapsto \mathbb{E}\left[\langle A, \log M_{J,k,n} - \mathbb{E}[\log M_{J,k,n}]\rangle_F \ (\log M_{J,k,n} - \mathbb{E}[\log M_{J,k,n}])\right]. \end{cases}$$

It is easy to see, that $\mathcal{C}(k2^{-J})$ is the unique symmetric positive definite operator such that, for all $A \in Sym(d)$,

$$\mathrm{Var}\,\langle \log M_{J,k,n},\ A\rangle_F = \left\langle \mathcal{C}(k2^{-J})A,\ A\right\rangle_F. \tag{3.7}$$

Note that the consistency condition (3.2) also implies $\mathcal{C}(k2^{-J}) = \mathcal{C}(k'2^{-J'})$ whenever $k2^{-J} = k'2^{-J'}$. Since the dyadic numbers are dense in $[0,1]$ and since our investigations require us to let J (as well as the threshold scale J_0) go to infinity, we make additionally the following (technical) assumption:

Assumption 3.4. The set of covariance operators $\{\mathcal{C}(t) \mid t \in [0,1] \text{ is dyadic}\}$ corresponding to the set of logarithmized scaling coefficients $\{\log M_{J,k,n} \mid k = 0,\dots,2^J-1 \text{ and } J \geq 0\}$ via

$$\mathcal{C}(t) = \mathrm{Cov}(\log M_{J,k,n}),\ \text{whenever } t = k2^{-J}$$

can be extended to a càdlàg process $(\mathcal{C}(t) : t \in [0,1])$, unique by construction, of symmetric positive definite operators $\mathcal{C}(t)\colon Sym(d) \to Sym(d)$, that is $\langle \mathcal{C}(t)A, A\rangle_F > 0$ for all $t \in [0,1]$ and $A \in Sym(d)$.

To illustrate our assumption on the covariance structure of the sampling scheme, let us consider the Riemannian signal plus noise model again:

Example 2. *In Example 1 we discussed the model* $M_{J,k,n} = \exp(\log(c_k) + \xi_k)$ *with i.i.d. noise* $(\xi_k)_k$ *and showed that Assumption 3.1 (r) is satisfied as long as* $\mathbb{E}\|\xi_k\|_F^r < \infty$ *for some* $r \geq 1$. *Let us now assume* $r \geq 2$.

It follows directly from the model that $\log M_{J,k,n} = \log(c_k) + \xi_k$ *and therefore*

$$\mathrm{Cov}(\log M_{J,k,n}) = \mathrm{Cov}(\xi_k) = \mathrm{Cov}(\xi_0),$$

which in turn implies, the extension of the covariance operators is trivially given by $\mathcal{C}(t) := \mathcal{C} = \mathrm{Cov}(\xi_0)$, $t \in [0,1]$. *Note that* $\mathcal{C} = \mathrm{Cov}(\xi_0)\colon Sym(d) \to Sym(d)$, $A \mapsto \mathbb{E}[\langle A, \xi_0\rangle_F \xi_0]$ *is a 4th order tensor and thus can be represented (with respect to the standard basis) by an array* $(C_{ijnm})_{i,j,n,m}$, *that is*

$$(\mathcal{C}A)_{i,j} = \sum_{n,m} C_{ijnm} A_{nm}, \quad A = (A_{nm})_{n,m} \in Sym(d). \tag{3.8}$$

Since the covariance operator $\mathcal{C}$ *is also uniquely characterized by* $\langle \mathcal{C}A, A\rangle_F = \mathrm{Var}\langle \xi_0, A\rangle_F$ *for all* $A \in Sym(d)$, *writing out equation (3.8) yields*

$$C_{ijnm} = \mathrm{Cov}\left((\xi_0)_{ij},\ (\xi_0)_{nm}\right). \tag{3.9}$$

Now suppose the noise matrices ξ_k have independent entries, that is

$$\xi_k = \sum_{1\le i\le j\le d} z_{ij}^k E_{ij},$$

where the $(z_{ij}^k)_{i,j,k}$ are independent with $\mathbb{E}z_{ij}^k = 0$ and $\mathrm{Var}(z_{ij}^k) = \sigma_{ij}^2$. Here, the matrices $E_{ij} \in Sym(d)$ are defined by

$$(E_{ij})_{k,l} = \begin{cases} 1 & , \quad i = k, j = l \quad or \quad i = l, j = k \\ 0 & , \quad else \end{cases}$$

and obviously form a basis for $Sym(d)$. Then in this specific case the covariance structure (3.9) reduces to

$$C_{ijnm} = \mathrm{Cov}\left(z_{ij}^k,\ z_{nm}^k\right) = \begin{cases} \sigma_{ij}^2 & , \quad i = n, j = m \quad or \quad i = m, j = n \\ 0 & , \quad else. \end{cases}$$

3.1. Asymptotic Normality of the AI wavelet estimator

We are now in position to present our main result on the asymptotic normality of the (logarithmized) linear wavelet estimator (3.6), that is, in the following we assume $g(\cdot) \equiv 0$ for the threshold function in (3.4). The weak limit of $\log \widehat{M}_{J,k,n}(J_0)$ will follow, on the one hand, from the weak limit of the underlying sampling scheme, i.e. the weak limit of $\log M_{J,k,n}$, combined with the convergence (2.15) of the AI-refinement scheme, on the other hand.

Let us start by noting that, for a fixed scale $j < J$, a recursive application of the midpoint pyramid algorithm (2.6) yields the representation

$$\log M_{j,k,n} = 2^{-(J-j)} \sum_{u=0}^{2^{J-j}-1} \log M_{J,u+k2^{J-j},n}. \tag{3.10}$$

Recall that the covariance operator $\mathrm{Cov}(\log M_{j,k,n}) = \mathcal{C}(k2^{-j})$ is uniquely characterized by (3.7) and therefore, due to the independence of the $(M_{J,k,n})_k$, we have by a Riemann approximation, for $A \in Sym(d)$,

$$\begin{aligned} 2^{J-j}\mathrm{Var}\left\langle \log M_{j,k,n}, A\right\rangle &= 2^{-(J-j)} \sum_{u=0}^{2^{J-j}-1} \mathrm{Var}\left\langle \log M_{J,u+k2^{J-j},n}, A\right\rangle \\ &= 2^{J-j}\left\langle \mathcal{C}(k2^{-j})A, A\right\rangle \\ &= 2^{-(J-j)} \sum_{u=0}^{2^{J-j}-1} \left\langle \mathcal{C}(u2^{-J} + k2^{-j})A, A\right\rangle \\ &\xrightarrow{J\to\infty} 2^j \int_{k2^{-j}}^{(k+1)2^{-j}} \langle \mathcal{C}(t)A, A\rangle \, \mathrm{d}t. \end{aligned} \tag{3.11}$$

Now, the uniform bounded moments condition (3.1) implies (see Lemma B.2), that the process $(\mathcal{C}(t) : t \in [0,1])$ in Assumption 3.4 is uniformly bounded as well, in the sense that

$$\sup_{t\in[0,1]} \|\mathcal{C}(t)\|_{\mathcal{L}} \leq \sup_{J,k} \mathbb{E}\|\log M_{J,k,n}\|_F^2 \leq K(r) < \infty. \tag{3.12}$$

Here $\mathcal{L} = \mathcal{L}(Sym(d))$ denotes the (Banach-) space of bounded linear operators $T\colon Sym(d) \to Sym(d)$ equipped with the usual induced operator norm $\|T\|_{\mathcal{L}} = \sup\{\|TA\|_F \mid A \in Sym(d),\ \|A\|_F \leq 1\}$. As immediate consequence, we have $\int_I \|\mathcal{C}(t)\|_{\mathcal{L}}\,\mathrm{d}t < \infty$ for any interval $I \subset [0,1]$. This in turn is a necessary and sufficient condition for existence of the Bochner integral $\int_I \mathcal{C}(t)\,\mathrm{d}t \in \mathcal{L}$, which is the unique element satisfying $\int_I \psi(\mathcal{C}(t))\,\mathrm{d}t = \psi\left(\int_I \mathcal{C}(t)\,\mathrm{d}t\right)$ for all linear and continuous functionals $\psi\colon \mathcal{L} \to \mathbb{R}$. Therefore, applied to the functional $\psi_A\colon T \mapsto \langle TA, A\rangle$, we obtain from (3.11), that

$$2^{J-j}\mathrm{Var}\,\langle \log M_{j,k,n}, A\rangle \quad \xrightarrow{J\to\infty} \quad \left\langle 2^j \int_{k2^{-j}}^{(k+1)2^{-j}} \mathcal{C}(t)\,\mathrm{d}t\ A, A \right\rangle. \tag{3.13}$$

Also, to give some further intuition on the limiting variance structure in our central limit theorems, observe that for small interval sizes, i.e. j large enough, we can approximate the RHS in (3.13) by $\langle \mathcal{C}(k2^{-j})A, A\rangle$ due to the assumed càdlàg property of $(C(t) : t \in [0,1])$.

The convergence (3.13) together with a verification of Lyapunov's condition are the main ingredients of the following proposition regarding the weak limit of the logarithmized empirical midpoints respectively scaling coefficients:

Proposition 3.5 (Asymptotic normality of $\log M_{j,k,n}$)**.** *Granted Assumption 3.1(r) and Assumption 3.4 are satisfied for some $r > 2$.*

(a) Let $0 < j < J$ and $0 \leq k \leq 2^j - 1$ be arbitrary but fixed. Then, as $n = 2^J \to \infty$,

$$2^{-j/2}n^{1/2}\Big(\log M_{j,k,n} - \mathbb{E}\big[\log M_{j,k,n}\big]\Big) \quad \Longrightarrow \quad \mathcal{N}\big(0, \Sigma_{j,k}\big),$$

in the Hilbert space $(Sym(d), \langle\cdot,\cdot\rangle)$, where the covariance operator $\Sigma_{j,k}$ is given by

$$\Sigma_{j,k} = 2^j \int_{k2^{-j}}^{(k+1)2^{-j}} \mathcal{C}(t)\,\mathrm{d}t.$$

(b) Let $x \in [0,1]$ be a continuity point of $\mathcal{C}$ and let $(J_{0,n})_n$, $(k_n)_n$ be sequences such that $J_{0,n} \to \infty$ and $J - J_{0,n} \to \infty$ and $k_n 2^{-J_{0,n}} \to x$. Then, as $n = 2^J \to \infty$,

$$2^{-J_0/2}n^{1/2}\Big(\log M_{J_0,k,n} - \mathbb{E}\big[\log M_{J_0,k,n}\big]\Big) \quad \Longrightarrow \quad \mathcal{N}\big(0, \mathcal{C}(x)\big).$$

Before we state our main result, let us note that the consistency of our proposed linear wavelet estimator has already been established in Chau [2018] and Chau and von Sachs [2021] for a general class of Riemannian metrics, including the considered log-Euclidean metric. The respective result can be summarized as follows: Under suitable regularity conditions on the underlying curve c, there is a constant $K > 0$ depending only on the refinement order N and thresholding scale J_0 such that

$$\mathbb{E}\left\|\log \widehat{M}_{J,k,n}(J_0) - \mathbb{E}\left[\log \widehat{M}_{J,k,n}(J_0)\right]\right\|_F^2 \leq K\, 2^{-(J-J_0)}.$$

A detailed version of the statement is provided in the Supplement, see Proposition B.1.

Deriving the weak limit of the linear wavelet estimator will additionally require to take into account the convergence (2.15) of the AI-refinement scheme. In particular, the limit

$$E_N^\infty := \lim_{m\to\infty} E_N^m = \Phi_L(0) \in \mathbb{R}^{(2N-1)\times(2N-1)}$$

will enter into the asymptotic variance in form of the quantity

$$\kappa_N := \sum_{i=1}^{2N-1} (E_N^\infty)_{1,i}^2 = \sum_{i=-2L}^{2L} \varphi_L^2(i). \tag{3.14}$$

A factor of the form κ_N is quite common to appear in the weak limit of wavelet estimators. To give an example, we point out that κ_N defined via (3.14) here, coincides with the term w_0^2 which appears in the variance of the Gaussian approximation in Theorem 3.4 in Antoniadis et al. [1994] for scalar-valued first generation wavelet estimators.

Theorem 3.6 (Asymptotic normality of the linear wavelet estimator)**.** *Let $x \in (0,1)$ be a* dyadic *continuity point of $\mathcal{C}$ and let $(J_0)_n = (J_{0,n})_n$, $(k)_n = (k_n)_n$ be sequences such that $J_{0,n} \to \infty$ and $J - J_{0,n} \to \infty$ and $k_n 2^{-J_{0,n}} \to x$. Granted Assumption 3.1(r) and Assumption 3.4 are satisfied for some $r > 2$, then the linear wavelet estimator $\widehat{M}_{J,k_n,n}(J_{0,n})$ obtained via (3.6) for a refinement order $N = 2L+1$ satisfies, as $n = 2^J \to \infty$,*

$$2^{-J_0/2} n^{1/2} \Big(\log \widehat{M}_{J,k,n}(J_0) - \mathbb{E}\big[\log \widehat{M}_{J,k,n}(J_0)\big] \Big) \quad \Longrightarrow \quad \mathcal{N}\big(0, \kappa_N\, \mathcal{C}(x)\big)\ .$$

Remark 3.7. Clearly, Theorem 3.6 is closely linked to Proposition 3.5 because the convergence to a Gaussian distribution is induced by the midpoint pyramid algorithm, which is applied in both statements. The main difference between both results consists in the final lifting step from the coarse scale J_0 back to the initial (sampling) scale J. This step rescales the limiting covariance operator from $\mathcal{C}(x)$ to $\kappa_N \mathcal{C}(x)$. Although we have not been able to prove, we conjecture that $\kappa_N < 1$ if $L = (N-1)/2 > 1$ which shows a decrease of the value of the asymptotic variance. With some effort we were able to analytically compute κ_N for $N = 1, 3, 5, 7$. The results are summarized in Table 1.

	$N=1$	$N=3$	$N=5$	$N=7$
κ_N	1	$\frac{25}{36} \approx 0.69$	$\frac{168549}{213160} \approx 0.79$	$\frac{107721892723}{126282847320} \approx 0.85$

TABLE 1
Different values of κ_N

As usual with asymptotics ($n \to \infty$) for estimating curves sampled on an asymptotically finer and finer grid, one has to understand the above given expression for the asymptotic variance $\kappa_N C(x)$ in the sense that the reference location x has to remain the *same* fixed dyadic rational in $(0,1)$ for all n. In general, i.e. for $x \in (0,1)$ not remaining the same fixed dyadic number, we are still be able to formulate a CLT of the following form:

Proposition 3.8. *Let $x \in (0,1)$ be a continuity point of $\mathcal{C}$ and let $(J_0) = (J_{0,n})_n$, $(k)_n = (k_n)_n$ be sequences such that $J_{0,n} \to \infty$ and $J - J_{0,n} \to \infty$ and $k_n 2^{-J_{0,n}} \to x$. Granted the conditions of Theorem 3.6 are satisfied, then the linear wavelet estimator $\widehat{M}_{J,k_n,n}(J_{0,n})$ obtained via (3.6) for a refinement order $N = 2L+1$ satisfies, as $n = 2^J \to \infty$,*

$$\mathrm{Var}\Big(\Big\langle \log \widehat{M}_{J,k,n}(J_0) - \mathbb{E}\big[\log \widehat{M}_{J,k,n}(J_0)\big],\ A\Big\rangle\Big)^{-1/2} \cdot$$
$$\cdot \Big\langle \log \widehat{M}_{J,k,n}(J_0) - \mathbb{E}\big[\log \widehat{M}_{J,k,n}(J_0)\big],\ A\Big\rangle \quad \Longrightarrow \quad \mathcal{N}(0,1).$$

Remark 3.9. We shed a bit more light onto the fact that for non-dyadic numbers no stable limiting variance of the wavelet estimator may exist, and that we can only give the limiting behavior of the wavelet estimator in the "standardized" version of Proposition 3.8. This phenomenon has already been observed in Antoniadis et al. [1994], Theorem 3.3 and its discussions, for ordinary scalar-valued wavelet estimators. Essentially it stems from the fact that for non-dyadic numbers on scale j of the form $t_j = 2^j t - [2^j t] \neq 0$, the sequence t_j - and hence the asymptotic variance expressed as a function of t_j - fails to converge. In our situation this can be translated to non-convergence of the matrix products of the type $U_J U_{J-1} \dots U_{J_0+1}$ with $U_i \in \{(E_N)_{\searrow d}, (O_N)_{\searrow d}\}$, as occurring in Lemma 2.2 respectively in (2.15), see also (B.4) in the Supplement. Recall that these products describe the lifting of the AI wavelet estimator smoothed on scale J_0 back to the sampling scale J. As both scales J_0 and J need to tend to infinity simultaneously (though on different rates), convergence to a finite limit depends on the fact whether there exists a n_0 beyond which the sequence $t_{n,0} := k_n 2^{-J_{0,n}} = x \in (0,1)$ for all $n > n_0$ (case of a dyadic point) or not: The affirmative case amounts to considering $J_0 = J_{0,n}$ (i.e. the element U_{J_0+1} in the matrix product) akin to remain fixed for all $n > n_0$, and the considered infinite matrix product possesses a limit which is given by (2.15). In the opposite case U_{J_0+1} continues to change with n, leading to no convergence of the matrix product. By the proof of Proposition 3.8 we can, however, show that the behaviour remains somewhat controlled in this latter case.

Remark 3.10. With our derived Central Limit Theorem 3.6 we observe strong parallels in the given rates with those from classical nonparametric curve es-

timation. To control variance and bias of our estimator on sample scale J, we need the usual conditions that $J_0 = J_{0,n} \to \infty$ and $J_n - J_{0,n} \to \infty$, for which the precise asymptotic rates can be obtained as usual by matching the (squared) asymptotic bias and variance as in Proposition B.1. Note, however, that this optimal choice $J_0 = \lfloor J/(2N+1) \rfloor$ is merely of asymptotic flavour, as in practice, for reasonably small sample sizes $n = 2^J$, this would lead to far too small scales J_0.

4. Confidence sets for SPD matrix-valued curves

In this section we derive confidence sets for SPD matrix-valued curves based on the Gaussian approximation of the linear wavelet estimator, see Theorem 3.6. Due to their asymptotic nature, these confidence sets require the sampling scale J and the threshold scale J_0 as well as their difference $J - J_0$ to be sufficiently large, in order to reach their aimed coverage level to a satisfactory extend. To circumvent typical issues that arise for small sample sizes $n = 2^J$, we also propose confidence sets based on a wild bootstrap scheme. With both approaches, in particular with the latter one, we introduce some interesting inference methods for non-scalar valued non-parametric curve estimators.

4.1. Asymptotic confidence sets

Let us briefly outline how the asymptotic normality in Theorem 3.6 leads to corresponding confidence sets of level $(1-\alpha)$ in a more or less canonical manner: To that end, we may first recall, that a random SPD matrix X is said to follow a *log-normal* distribution (w.r.t. the log-Euclidean metric) with mean $S \in Sym^+(d)$ and covariance operator $\Sigma \in \mathcal{L}(Sym(d))$ if and only if

$$\langle \log X, A \rangle_F \sim \mathcal{N}(\langle \log S, A \rangle_F, \langle \Sigma A, A \rangle_F) \quad \forall A \in Sym(d). \tag{4.1}$$

Now a straightforward way to obtain distributional properties for log-normal random matrices is to utilize that $Sym(d)$ and $\mathbb{R}^q$, $q = d(d+1)/2$, are isometrically isomorphic. To elaborate, consider the mapping $\eta\colon Sym(d) \to \mathbb{R}^q$ defined by,

$$\eta(A) := (A_{11}, \dots, A_{dd}, \sqrt{2}A_{12}, \dots, \sqrt{2}A_{1d}, \sqrt{2}A_{23}, \dots, \sqrt{2}A_{2d}, \dots, \sqrt{2}A_{d-1,d})^T, \quad A = (A_{ij})_{ij},$$

i.e. η stacks first the diagonal elements and then row-wise the above diagonal elements (scaled by $\sqrt{2}$) of a symmetric matrix into a vector. It is easy to see that η is one-to-one and satisfies $\langle \eta A, \eta B \rangle = \langle A, B \rangle_F$. Thus, η is a linear isometry between $Sym(d)$ and $\mathbb{R}^q$. Moreover, a random SPD matrix X has a log-normal distribution in the sense of (4.1) if and only if

$$\eta(\log X) \sim \mathcal{N}(\eta(\log S), \Sigma^\eta) \quad \text{with} \quad \Sigma^\eta := \mathrm{Cov}(\eta(\log X)) = \eta\, \Sigma\, \eta^{-1}.$$

Applying this insight to Theorem 3.6, granted the assumptions are satisfied, immediately shows

$$2^{-J_0/2}n^{1/2}\ \eta\left(\log \widehat{M}_{J,k,n}(J_0) - \mathbb{E}\big[\log \widehat{M}_{J,k,n}(J_0)\big]\right) \quad \Longrightarrow \quad \mathcal{N}\big(0, \kappa_N\ \mathcal{C}(x)^\eta\big), \tag{4.2}$$

where $x = k^* 2^{-J^*} \in (0,1)$ is a dyadic rational and $\mathcal{C}(x)^\eta = \eta\,\mathcal{C}(x)\,\eta^{-1}$. Suppose $\widehat{\mathcal{C}}(x)^\eta$ is a consistent estimator for the covariance matrix $\mathcal{C}(x)^\eta$. Then, as $(\kappa_N\ \mathcal{C}(x)^\eta)^{-1/2}\mathcal{N}(0, \kappa_N\ \mathcal{C}(x)^\eta)$ agrees with $\mathcal{N}(0, \mathrm{Id}_{q\times q})$, we can conclude with Slutsky's theorem and the continuous mapping theorem that

$$\begin{aligned}
&2^{-J_0}n\ \left\|(\kappa_N\ \widehat{\mathcal{C}}(x)^\eta)^{-1/2}\eta\left(\log \widehat{M}_{J,k,n}(J_0) - \mathbb{E}\big[\log \widehat{M}_{J,k,n}(J_0)\big]\right)\right\|^2 \\
&= 2^{-J_0}n\ \eta\left(\log \widehat{M}_{J,k,n}(J_0) - \mathbb{E}\big[\log \widehat{M}_{J,k,n}(J_0)\big]\right)^T \left(\kappa_N\ \widehat{\mathcal{C}}(x)^\eta\right)^{-1} \eta\left(\log \widehat{M}_{J,k,n}(J_0) - \mathbb{E}\big[\log \widehat{M}_{J,k,n}(J_0)\big]\right) \\
&\Rightarrow \chi^2_q,
\end{aligned} \tag{4.3}$$

where χ^2_q denotes the chi-squared distribution with q degrees of freedom. Therefore, if we let $\chi^2_{q,1-\alpha}$ denote the $(1-\alpha)$-quantile of the χ^2_q-distribution, it follows under the conditions of Theorem 3.6, that

$$CS_{n,1-\alpha}\left(\mathbb{E}\left[\log \widehat{M}_{J^*,k^*}\right]\right) := \left\{ S \in Sym(d) \;\middle|\; 2^{-J_0} n\,\kappa_N^{-1} \left\|\left(\widehat{\mathcal{C}}(x)^\eta\right)^{-1/2} \eta\left(\log \widehat{M}_{J,k,n}(J_0) - S\right)\right\|^2 \le \chi^2_{q,1-\alpha} \right\}, \tag{4.4}$$

is an asymptotic $(1-\alpha)$-confidence set for $\mathbb{E}[\log \widehat{M}_{J^*,k^*}] := \lim_{n\to\infty} \mathbb{E}[\log \widehat{M}_{J,k,n}(J_0)] \in Sym(d)$.

Furthermore, regarding the bias of the linear estimator, we have by Proposition B.1, that

$$2^{-J_0/2}n^{1/2}\left\|\mathbb{E}\big[\log \widehat{M}_{J,k,n}(J_0)\big] - \log M_{J^*,k^*}\right\|_F \le C 2^{J-(2N+1)J_0/2}, \tag{4.5}$$

if the target signal $c\colon [0,1] \to Sym^+(d)$ is N-times (covariant) differentiable with respect to the log-Euclidean metric. Thus, choosing J and J_0 such that

$$J, J_0 \to \infty, \quad J - (2N+1)J_0 \to -\infty, \quad J - J_0 \to \infty$$

yields on the one hand

$$2^{-J_0/2}n^{1/2}\left\|\eta\left(\mathbb{E}\big[\log \widehat{M}_{J,k,n}(J_0)\big] - \log M_{J^*,k^*}\right)\right\| \le C 2^{J-(2N+1)J_0/2} \to 0,$$

and on the other hand the weak convergence (4.2) still holds true. Therefore, by Slutsky's theorem again,

$$2^{-J_0/2}n^{1/2}\ \eta\left(\log \widehat{M}_{J,k,n}(J_0) - \log M_{J^*,k^*}\right)$$

$$= 2^{-J_0/2} n^{1/2}\ \eta\left(\log \widehat{M}_{J,k,n}(J_0) - \mathbb{E}\big[\log \widehat{M}_{J,k,n}(J_0)\big]\right) + 2^{-J_0/2} n^{1/2}\ \eta\left(\mathbb{E}\big[\log \widehat{M}_{J,k,n}(J_0)\big] - \log M_{J^*,k^*}\right)$$
$$\Longrightarrow \quad \mathcal{N}(0, \kappa_N\ \mathcal{C}(x)^\eta) \tag{4.6}$$

and with the same reasoning as above, we see that

$$CS_{n,1-\alpha}\left(M_{J*,k*}\right) := \left\{ S \in Sym^+(d) \;\middle|\; 2^{-J_0} n\, \kappa_N^{-1} \left\| \left(\widehat{\mathcal{C}^\eta}(x)\right)^{-1/2}\ \eta\left(\log \widehat{M}_{J,k,n}(J_0) - \log S\right)\right\|^2 \leq \chi^2_{q,1-\alpha} \right\} \tag{4.7}$$

poses an asymptotic $(1-\alpha)$-confidence set for $M_{J^*,k^*} \in Sym^+(d)$, granted the target c is N-times differentiable, the conditions of Theorem 3.6 are satisfied and $J - (2N+1)J_0 \to -\infty$.

Remark 4.1. The additional requirement $J-(2N+1)J_0 \to -\infty$ on the threshold scale J_0 is commonly referred to as *undersmoothing*. It is necessary to compensate the bias in the weak limit of the linear estimator. Note that the optimal choice, $J_0 = J/(2N+1)$, which simultaneously minimizes bias and variance (see Proposition B.1), is right at the edge of undersmoothing, since in this case $J - (2N+1)J_0$ is constant.

Remark 4.2. If the difference $J - J_0$ is not substantial enough, the Gaussian approximation (4.2) respectively (4.6) may of course not be a sufficiently reasonable choice to mimic the distribution of $2^{-J_0/2} n^{1/2}\, \eta(\log \widehat{M}_{J,k,n}(J_0) - \log M_{J^*,k^*})$. Moreover, the assumed consistent estimator $\widehat{\mathcal{C}}(x)^\eta$ for the covariance $\mathcal{C}(x)^\eta = \eta\, \mathcal{C}(x)\, \eta^{-1}$ can converge rather slowly in practice, so that the corresponding χ^2-approximation (4.3) may additionally suffer if the sample size $n = 2^J$ is not adequate. For these reasons the asymptotic confidence sets (4.4) and (4.7) are expected to reach their coverage to a satisfactory degree only if both the sampling scale J and the difference $J - J_0$ are sufficiently large.

To illustrate the asymptotic confidence sets (4.7), we continue our discussion of the intrinsic Riemannian signal plus noise model:

Example 3. *We have already shown in Example 1 and 2 that the signal plus noise model $M_{J,k,n} = \exp\{\log(c_k) + \xi_k\}$, with $c_k = c(k/n)$ and i.i.d. noise $(\xi)_k$ such that $\mathbb{E}\|\xi_k\|_F^r < \infty$ for some $r > 2$, satisfies Assumption 3.1(r) and Assumption 3.4. Moreover, recall that the covariance operator in this model is constant, i.e. $\mathcal{C}(x) = \mathcal{C} = \mathrm{Cov}(\xi_0)$ for $x \in [0,1]$, and in case $\xi_k = \sum_{1\leq i\leq j\leq d} z_{ij}^k E_{ij}$ has independent entries $(z_{ij}^k)_{i,j,k}$ with $\mathbb{E}[z_{ij}^k] = 0$ and $\mathrm{Var}(z_{ij}^k) = \sigma_{ij}^2$, the coefficients of $\mathcal{C}$ (with respect to the standard basis) are just*

$$C_{ijnm} = \begin{cases} \sigma_{ij}^2, & i = n, j = m \text{ or } i = m, j = n \\ 0, & \text{otherwise.} \end{cases}$$

Thus, a straightforward calculation reveals

$$\mathcal{C}^\eta = \eta\ \mathcal{C}\ \eta^{-1} = \mathrm{diag}\left(\sigma_{11}^2, \ldots, \sigma_{dd}^2, 2\sigma_{12}^2, \ldots, 2\sigma_{1d}^2, 2\sigma_{23}^2, \ldots, 2\sigma_{2d}^2, \ldots, 2\sigma_{d-1,d}^2\right).$$

Suppose $\widehat{\mathcal{C}}^\eta = \operatorname{diag}(\widehat{\sigma}^2_{11}, \dots, 2\widehat{\sigma}^2_{d-1,d})$ *is a consistent estimator for* $\mathcal{C}^\eta$ *and let* $m^{(n)}_{ij} = (\log \widehat{M}_{J,k,n}(J_0))_{ij}$ *as well as* $a_{ij} = (\log S)_{ij}$. *Then, the inequality in* (4.7) *translates to*

$$2^{J-J_0}\,\kappa_N^{-1} \sum_{1\le i\le j\le d} \frac{\left(m^{(n)}_{ij} - a_{ij}\right)^2}{\widehat{\sigma}^2_{ij}} \le \chi^2_{q,1-\alpha}.$$

Viewed as an inequality in the entries $(a_{ij})_{ij}$, *this subset of* $Sym(d)$ *corresponds via the isometry* η *to an ellipsoid in* $\mathbb{R}^q$ *with center* $\left(m^{(n)}_{11}, \dots, m^{(n)}_{dd}, \sqrt{2}m^{(n)}_{12}, \dots, \sqrt{2}m^{(n)}_{d-1,d}\right)$ *and half-axis*

$$\tau^{(n)}_{ij} = 2^{-(J-J_0)/2}\sqrt{\kappa_N\ \widehat{\sigma}^2_{ij}\ \chi^2_{q,1-\alpha}}, \quad 1 \le i \le j \le d.$$

Let $\mathcal{E}\left((m^{(n)}_{ij}, \tau^{(n)}_{ij})_{ij}, N, q, \alpha\right)$ *denote this ellipsoid, then we obtain for the confidence set* (4.7) *the compact representation*

$$CS_{n,1-\alpha}\left(M_{J^*,k^*}\right) = \left\{\exp(A) \;\middle|\; A = \eta^{-1}(x) \;\; \textit{for} \;\; x \in \mathcal{E}\left((m^{(n)}_{ij}, \tau^{(n)}_{ij})_{ij}, N, q, \alpha\right)\right\}. \tag{4.8}$$

To demonstrate, consider the case $d = 2$. *Then* $q = 3$ *and* (4.8) *reduces to*

$$CS_{n,1-\alpha}\left(M_{J^*,k^*}\right) = \left\{ \exp\left(\begin{bmatrix} a_{11} & a_{12}/\sqrt{2} \\ a_{12}/\sqrt{2} & a_{22} \end{bmatrix} \right) \;\middle|\; \begin{pmatrix} a_{11} \\ a_{22} \\ a_{12} \end{pmatrix} \in \mathcal{E}\left((m^{(n)}_{ij}, \tau^{(n)}_{ij})_{ij}, N, 3, \alpha\right) \right\}.$$

4.2. Wild bootstrap confidence sets

In view of the common disadvantages of confidence sets based on asymptotic approximations (see Remark 4.2) for small sample sizes, we also propose confidence sets based on bootstrap approximations. The resampling procedure works as follows:

The bootstrap procedure. Let $X = (X_k)_{k=0,\dots,n-1}$ be a sequence of SPD matrices sampled on a dyadic scale J such that $2^J = n$ and let J_0^* denote a bootstrap threshold parameter. Similar to J_0, the threshold J_0^* is assumed to be asymptotically coarser than J in the sense that both $J - J_0^* \to \infty$ and $J_0^* \to \infty$ as $n \to \infty$.

Step 1. Set $(M_{J,k,n})_k = X_k$ for $k = 0, \dots, n-1$ and compute the linear wavelet estimator $(\widehat{M}_{J,k,n}(J_0^*))_k$ according to (3.6) w.r.t. the threshold level J_0^* and a refinement order $N = 2L+1$, $L \ge 0$.

Step 2. Calculate the residuals $\widehat{\varepsilon}_{J,k,n}(J_0^*) = \log \widehat{M}_{J,k,n}(J_0^*) - \log M_{J,k,n}$ for $k = 0, \dots, n-1$.

Step 3. Simulate bootstrap residuals

$$\varepsilon^*_{J,k,n}(J_0^*) = \widehat{\varepsilon}_{J,k,n}(J_0^*)\, V_{J,k,n} \quad \text{for} \quad k = 0, \dots, n-1,$$

where the $V_{J,k,n}$ are i.i.d., real-valued random variables with mean zero, unit variance and finite $(2+\delta)$th moment for some $\delta > 0$.

Step 4. Create bootstrap observations $M^*_{J,k,n}(J_0^*) = \exp\Big(\log \widehat{M}_{J,k,n}(J_0^*) + \varepsilon^*_{J,k,n}(J_0^*)\Big)$ for $k = 0, \dots, n-1$.

Step 5a. Compute the midpoint pyramid $(M^*_{j,k,n}(J_0^*))_{j,k}$ based on the bootstrap observations $(M^*_{J,k,n}(J_0^*))_k$.

Step 5b. Based on $(M^*_{J,k,n}(J_0^*))_k$ compute the *bootstrap linear wavelet estimator* $(\widehat{M}^*_{J,k,n}(J_0))_k$ according to (3.6) w.r.t. the threshold level J_0 and the refinement order N in **Step 1**.

Now, let us consider

$$\mathfrak{M}_{J,k,n}(J_0) := 2^{-J_0/2} n^{1/2} \left(\log \widehat{M}_{J,k,n}(J_0) - \mathbb{E}\left[\log \widehat{M}_{J,k,n}(J_0)\right] \right),$$

as well as its bootstrap counterpart

$$\mathfrak{M}^*_{J,k,n}(J_0) := 2^{-J_0/2} n^{1/2} \left(\log \widehat{M}^*_{J,k,n}(J_0) - \mathbb{E}\left[\log \widehat{M}^*_{J,k,n}(J_0)\right] \right).$$

Then we have the following fundamental result.

Theorem 4.3 (Validity of the bootstrap)**.** *Let $x \in (0,1)$ be a* dyadic *continuity point of $\mathcal{C}$ and let $(J_0)_n = (J_{0,n})_n$, $(J_0^*)_n = (J^*_{0,n})_n$, $(k)_n = (k_n)_n$ be sequences such that $J_{0,n}$, $J^*_{0,n} \to \infty$ and $k_n 2^{-J_{0,n}}$, $k_n 2^{-J^*_{0,n}} \to x$ as well as*

$$\min\left\{(1 - J_0^*/J)\frac{r}{2} - (1 - J_0/J),\ (1 - J_0/J)\frac{r}{4}\right\} \geq (1+\delta) \quad \forall n \in \mathbb{N}$$

for some $\delta > 0$. Granted Assumption 3.1(r) and Assumption 3.4 are satisfied for some $r > 4$. Then, with probability 1*,*

$$\mathcal{L}^*\big(\mathfrak{M}^*_{J,k,n}(J_0)\big) \quad \Longrightarrow \quad \mathcal{N}\big(0, \kappa_N\, \mathcal{C}(x)\big), \quad n \to \infty,$$

conditionally on the data $\{M_{J,k,n} \mid k = 0, \dots, n-1\}$. In particular,

$$\sup_{x \in Sym(d)} \left| \mathbb{P}(\mathfrak{M}_{J,k,n}(J_0) \leq x) - \mathbb{P}^*(\mathfrak{M}^*_{J,k,n}(J_0) \leq x) \right| \to 0 \quad a.s., \quad n \to \infty$$

conditionally on the data. Here, for a random matrix X with values in $Sym(d)$, we define the distribution function by $\mathbb{P}(X \leq x) = \mathbb{P}(X_{i,j} \leq x_{i,j},\ \forall\ 1 \leq i,j \leq d)$ for $x \in Sym(d)$.

Note that we need the moment condition (3.1) to be satisfied for some $r > 4$ because this time we need in each setting a law of large numbers to hold

for the empirical variance. Regarding the base level J_0, we observe that given $r > 4(2N+1)/(2N)$, we can choose the optimal $J_0 = \lfloor J/(2N+1) \rfloor$ as well as $J_0^* = \lfloor \alpha J \rfloor$ for

$$\alpha \in \Big(0, 1 - \frac{2}{r}\big(1 + \frac{2N}{2N+1}\big)\Big).$$

Note that the choice $\alpha = 1/(2N+1)$ is admissible in this case. However, given these choices, we have to correct for the bias $2^{(J-J_0)/2}(\mathbb{E}[\log \widehat{M}_{J,k,n}(J_0)] - \log M_{J^*,k^*})$, $x = k^* 2^{-J^*}$, which does not vanish in the limit given these choices.

Remark 4.4. Concerning the choice of the base level J_0 we remark that when compared to the optimal choice $\lfloor J/(2N+1) \rfloor$ we distinguish two scenarios which already arise in bootstrap for traditional kernel regression. If $J_0 - \lfloor J/(2N+1) \rfloor \to -\infty$, we oversmooth the estimator; conversely, if $J_0 - \lfloor J/(2N+1) \rfloor \to \infty$, we undersmooth. Oversmoothing usually requires implicit or explicit bias correction as commonly suggested in the literature (see, e.g., Härdle and Bowman [1988]). In order to avoid this need for bias correction undersmoothing is recommended (as in Hall [1992], e.g.). In our present setting, including for the simulations of Section 5, undersmoothing is also favorable as the optimal choice $J_0 = \lfloor J/(2N+1) \rfloor$ would be far too small even for moderate scales J.

Based on the validity result in Theorem 4.3 we propose as bootstrap confidence sets

$$\begin{aligned} CS^*_{n,1-\alpha}(M_{J^*,k^*}) &:= \Big\{ S \in Sym^+(d) \;\Big|\; d\Big(\widehat{M}_{J,k,n}(J_0), S\Big) < r_{\alpha,k} \Big\} \\ &= \exp\Big(\Big\{ A \in Sym(d) \;\Big|\; \big\|\eta\big(\log \widehat{M}_{J,k,n}(J_0)\big) - \eta(A)\big\| < r_{\alpha,k} \Big\}\Big) \\ &= \exp\Big(\eta^{-1}\Big(\Big\{ x \in \mathbb{R}^q \;\Big|\; \big\|\eta\big(\log \widehat{M}_{J,k,n}(J_0)\big) - x\big\| < r_{\alpha,k} \Big\}\Big)\Big), \end{aligned} \tag{4.9}$$

i.e. the metric balls with centre $\widehat{M}_{J,k,n}(J_0)$ and radius $r_{\alpha,k}$ with respect to the log-Euclidean metric. The radii $r_{\alpha,k}$, $k = 0, \dots, n$, are chosen respectively as the (empirical) $(1-\alpha)$-quantile of the distances

$$d\Big(\widehat{M}_{J,k,n}(J_0),\ \widehat{M}^{*,b}_{J,k,n}(J_0)\Big) = \Big\|\log \widehat{M}_{J,k,n}(J_0) - \log \widehat{M}^{*,b}_{J,k,n}(J_0)\Big\|_F, \qquad b = 1, \dots, B,$$

where B is the number of bootstraps iterations, i.e. the number of independent bootstrap estimators $(\widehat{M}^{*,b}_{J,k,n}(J_0))_k$ generated from the procedure at the beginning of this section.

5. Numerical simulation examples

In this simulation study we test our proposed estimation procedure for two different data-generating models. The first one is the already discussed signal plus noise model $M_{J,k,n} = \exp\{\log(c_k) + \xi_k\}$, where the noise matrices ξ_k have independent Gaussian entries $z^k_{ij} \sim \mathcal{N}(0, \sigma^2_{ij})$, see the continued example 1, 2

and 3 for a detailed discussion of the model. As target signal in this model we consider the following smooth $Sym^+(2)$-valued curves, for $t \in [0,1]$ (see Figure 3):

$$c_1(t) = \begin{pmatrix} 50\sqrt{1-(2t-1)^2}+0.1 & 2\sin(17\pi t) \\ 2\sin(17\pi t) & 50t+1 \end{pmatrix},$$
$$c_2(t) = \begin{pmatrix} 55\cos(5\pi(t+0.1)/11) & 50\sqrt{\sin(10\pi(t+0.1)/11)/2} \\ 50\sqrt{\sin(10\pi(t+0.1)/11)/2} & 55\sin(5\pi(t+0.1)/11) \end{pmatrix},$$
$$c_3(t) = \begin{pmatrix} 2(5-10t)^2 & 5-10t \\ 5-10t & 1 \end{pmatrix}.$$

For the different curves we choose the following variance parameters

$$\begin{aligned} c_1: &\quad \sigma_{11}=0.05, \quad \sigma_{22}=0.1, \quad \sigma_{12}=0.01 \\ c_2: &\quad \sigma_{11}=0.1, \quad \sigma_{22}=0.05, \quad \sigma_{12}=0.1 \\ c_3: &\quad \sigma_{11}=0.1, \quad \sigma_{22}=0.1, \quad \sigma_{12}=0.1 \end{aligned}$$

Additionally we also consider a second data-generating mechanism via the multiplicative eigenvalue model: In this model the target signal is of the form $c(t) = U(t)D(t)U(t)^T$, $t \in [0,1]$, where $U\colon [0,1] \to M(d)$ are orthogonal matrices (i.e. rotations) and $D(t) = \mathrm{diag}(\lambda_1(t), \ldots, \lambda_d(t))$ is diagonal with smooth functions $\lambda_i\colon [0,1] \to (0,\infty)$. By construction $c(t)$ is symmetric and has positive eigenvalues $\lambda_i(t)$, hence c is a curve in $Sym^+(d)$. We assume the noise acts multiplicatively on the eigenvalues, i.e. we observe $M_{J,k,n} = U(k/n)(D(k/n)\xi_k)U(k/n)^T$ for $k = 0, \ldots, n$, where $\xi_k = \mathrm{diag}(\xi_1^{(k)}, \ldots, \xi_d^{(k)})$ are independent diagonal noise matrices and $\xi_i^{(k)}$ are positive random variables with finite second moment. The model is related to the signal plus noise model since an application of the matrix logarithm (and its basic properties) yields $\log(M_{J,k,n}) = \log(c(k/n)) + U(k/n)\log(\xi_k)U(k/n)^T$. For our simulation we choose the following specific curve (see again Figure 3) $c_4(t) = U(t)\mathrm{diag}(\lambda_1(t), \lambda_2(t))U(t)^T$ with

$$\lambda_1(t) = 10 + 50\sin^2(2\pi t), \qquad \lambda_2(t) = 10 + 50\cos^2(2\pi t)$$
$$U(t) = \begin{bmatrix} \cos(\pi t/2) & -\sin(\pi t/2) \\ \sin(\pi t/2) & \cos(\pi t/2) \end{bmatrix}$$

and for the noise we choose $\xi_k = \mathrm{diag}(\varepsilon_1^{(k)}, \varepsilon_2^{(k)})$, where $\varepsilon_i^{(k)} \sim \mathrm{Uniform}[0.8, 1.2]$ are independent with respect to k and to i.

Our AI-wavelet estimators are depicted here for curves c_1 and c_4 in Figures 4 and 5, respectively, whereas the ones for c_2 can be found in Figure 7 and Figure 8 in the Supplement Section D.

The different parameters of these estimators are chosen to be $N = 5$, i.e. a moderate order of the AI-scheme, the sample scales vary between $J = 10$ and $J = 12$ (some analysis for the smaller sample size $n = 1024$ to be found in

Section D), and the smoothing scale $J_0 < J$ to be the remaining free parameter to be chosen curve by curve. For all simulations being based on $B = 100$ bootstrap repetitions (for which we choose $J_0^* = J_0$), we compute first the linear wavelet estimator $(\widehat{M}_{J,k,n}(J_0))_k$ according to (3.6) and then the corresponding asymptotic confidence sets (4.7) (respectively (4.8)) and bootstrap confidence sets (4.9) - as derived in the previous section - where we use standard normal $V_{J,k,n}$ in Step 3. In order to minimize additional variability due to estimation of the variance parameter, we replace the estimate $\widehat{\mathcal{C}}(x)^\eta$ by the true $\mathcal{C}(x)^\eta$, which is known in our simulation set-up. In order to visualize the results we identify positive definite matrices

$$M = \begin{pmatrix} x & z \\ z & y \end{pmatrix}$$

with points (x, y, z) in the cone $\mathfrak{C} = \{(x, y, z) \in \mathbb{R}^3 \mid xy > z^2\}$. Note that this correspondence is one-to-one. Both confidence sets are compared with respect to their empirical coverage based on $K = 100$ samples and, since the confidence sets can be considered as subsets in the cone $\mathfrak{C}$, we also approximate their respective (Lebesgue-) volumes. To compensate for the dependency of the volumes on the location within the cone $\mathfrak{C}$ - close to the boundary for example the confidence sets are flattened out, and they generally grow larger with increasing distance from the origin - we scale them by the volume that corresponds to the matrix exponential of the unit ball at the respective points.

In the following tables and figures, both for confidence regions based on asymptotic normality and based on our bootstrap, we report empirical coverages, first on average over all considered time points for curves $c_1(t), c_2(t)$ and $c_3(t)$ and for three different values of the nominal level. Furthermore, point by point plots of empirical coverages (omitting boundary points which suffer from the usual problems of nonparametric curve estimation) as well as the corresponding histograms are presented for curve $c_1(t)$, whereas those for $c_2(t)$ and $c_3(t)$ are reported in the Supplement Section D. Generally speaking our simulations show that the nominal levels are well respected. The resulting confidence regions are depicted, for curve $c_1(t)$, in Figure 4, and for curve $c_4(t)$ in Figure 5 whereas the plots for the two other curves are to be found again in the Supplement Section D.

Our simulations also showed that at certain points (e.g. of high curvature of the underlying target curve) the wavelet estimator suffers from the classical bias phenomenon of non-parametric curve estimators - in particular at the boundaries. We therefore omit the first and last 100 values in our computations. To illustrate this effect, more prominent for sample sizes of order 10^3 (i.e. for the choice of $J = 10$), we refer to Tables 5 and 6 in Section D where we compare for curve $c_1(t)$ the average empirical coverages for confidence regions centred around the target curve with those centred around the mean of the wavelet estimator which is approximated - at each considered point in time - by the sample mean over 50 repetitions.

TABLE 2
Simulation results for curve c_1 and $J = 12$, $J_0 = 7$, $N = 5$. Scaled averaged volumes are given in units of 10^{-4}.

	alpha=0.9		alpha=0.95		alpha=0.975	
	Asym.	Boot.	Asym.	Boot.	Asym.	Boot.
Emp. Cov.	0.8986	0.8701	0.9440	0.9252	0.9700	0.9542
Scaled Ave. Vol.	0.019	0.073	0.027	0.114	0.035	0.160

TABLE 3
Simulation results for curve c_2 and $J = 12$, $J_0 = 7$, $N = 5$. Scaled averaged volumes are given in units of 10^{-4}.

	alpha=0.9		alpha=0.95		alpha=0.975	
	Asym.	Boot.	Asym.	Boot.	Asym.	Boot.
Emp. Cov.	0.9084	0.8790	0.9498	0.9347	0.9714	0.9617
Scaled Ave. Vol.	0.155	0.216	0.217	0.321	0.284	0.434

TABLE 4
Simulation results for curve c_3 and $J = 12$, $J_0 = 7$, $N = 5$. Scaled averaged volumes are given in units of 10^{-4}.

	alpha=0.9		alpha=0.95		alpha=0.975	
	Asym.	Boot.	Asym.	Boot.	Asym.	Boot.
Emp. Cov.	0.9027	0.8822	0.9473	0.9364	0.9702	0.9625
Scaled Ave. Vol.	0.356	0.330	0.497	0.478	0.651	0.635

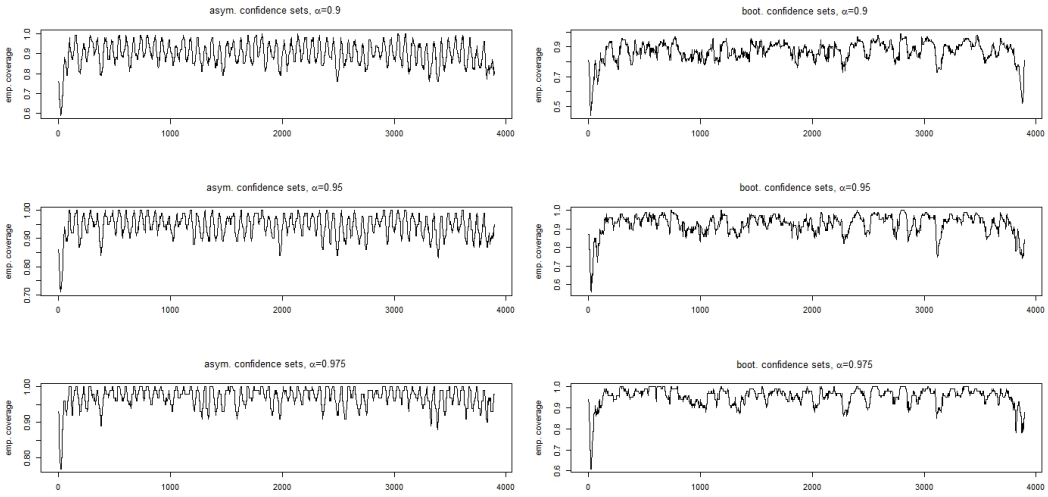


FIG 1. *Pointwise empirical coverage for curve* c_1 *and* $J = 12$, $J_0 = 7$, $N = 5$

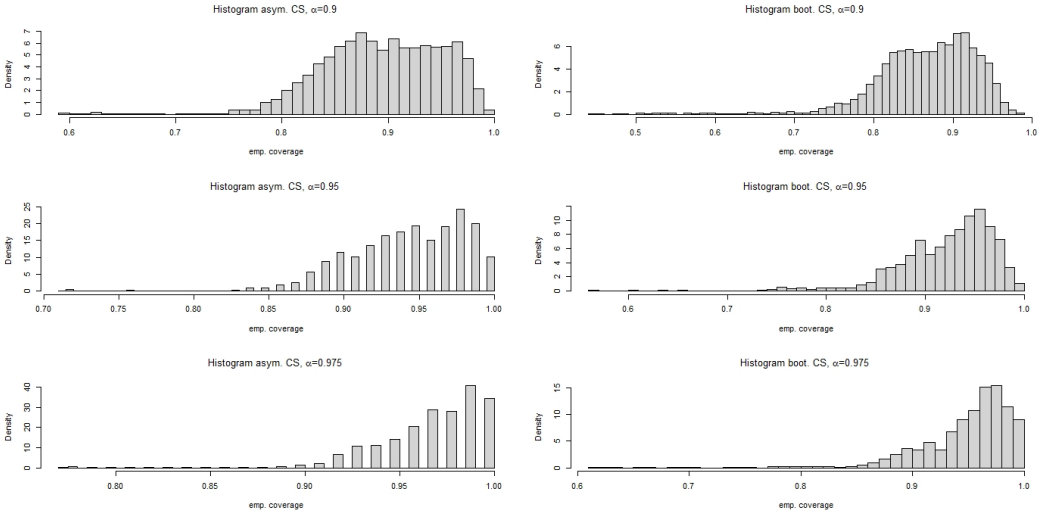


FIG 2. *Histogram of empirical coverage for curve* c_1 *and* $J = 12$, $J_0 = 7$, $N = 5$

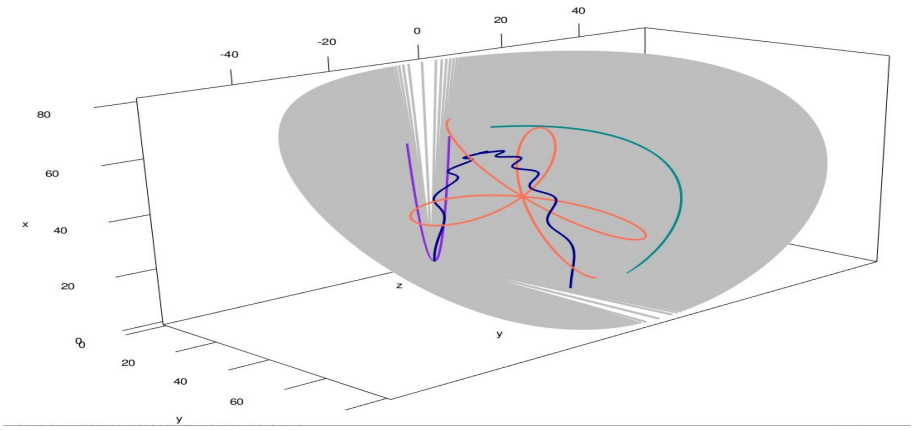


FIG 3. *True test curves:* c_1 = *blue,* c_2 = *green,* c_3 = *violet,* c_4 = *orange*

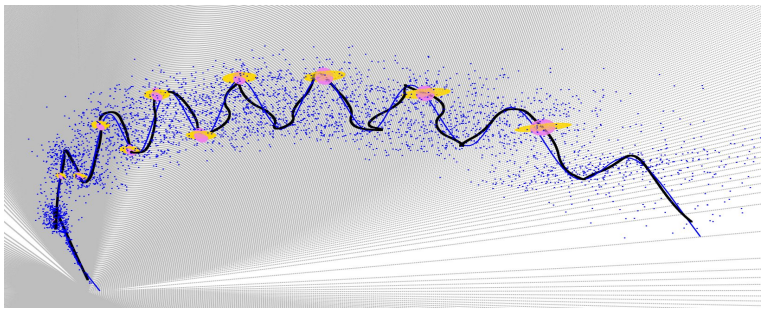


FIG 4. *AI-wavelet estimator (black), asymptotic (gold) and bootstrap (pink)* 0.975*-confidence sets for curve* c_1 *and* $J = 12$, $J_0 = 7$.

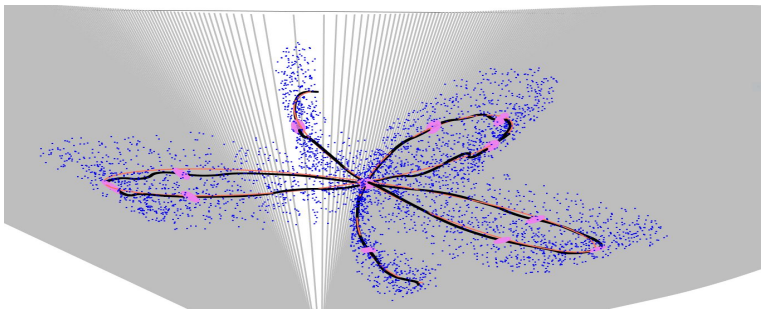


FIG 5. *AI-wavelet estimator (black) and bootstrap (pink)* 0.9*-confidence sets for curve* c_4 *and* $J = 12$, $J_0 = 7$.

Overall, we conclude that both our pointwise bootstrap-based confidence sets and the asymptotic ones work well in respecting the nominal levels. Not only our empirical coverages get close to the nominal ones on average (over all considered time points outside the boundaries) but we also observe only a small proportion where this is not the case. This is likely to be the case due to the usual potential bias problem (as discussed above) which is inherent to any nonparametric curve estimation procedure - and not specific to our more challenging matrix-valued estimation problem. Moreover, the closeness of both types of confidence sets supports the shown validity of our bootstrap method. Finally, our simulations show that in practice working with our bootstrap-based sets offers a computationally lighter method as it avoids needing to estimate the parameter $\mathcal{C}(x)^\eta$ arising in the asymptotic variances.

6. Conclusions

In this paper we have been proposing two constructions of (pointwise) confidence sets for SPD-matrix valued curves on a Riemannian manifold, based on (linear) AI wavelet estimators and the use of the log-Euclidean metric. The proven Central Limit Theorem enables us to construct asymptotic confidence regions. As alternative, we devel- oped (wild) bootstrap confidence schemes which we showed to be theoretically valid, empirically very satisfying and computationally easier to implement.

Quite naturally, the question arises about inference for the non-linear version of our wavelet estimator, as proposed by Chau and von Sachs [2021], based on trace-thresholding of the matrix-valued empirical wavelet coefficients. Constructing confidence sets for non-linear wavelet estimators is already a non-trivial task in the classical Euclidean set-up of first-generation wavelets for scalar-valued curves, the related literature is sparse (e.g., Chau and von Sachs [2016]). So although we believe that our approach based on the bootstrap sets can be successfully applied to the threshold wavelet estimator, we decided to leave this and further investigations on this question for future work. Note, however, that as a first theoretical step into this direction, we were able to derive a CLT, analogous to our Theorem 3.5, for the (whitened) empirical wavelet coefficients (2.10).

For the construction of our wavelet estimators, we used the log-Euclidean metric which is among the computationally easiest constructions of Riemannian metrics for the space of positive-definite symmetric matrices. Moreover, it enjoys important properties: it avoids swelling effects (very undesirable for confidence regions) and is invariant with respect to unitarian congruent transformation, allowing for permutation-equivariant estimators of covariance matrices of multivariate or time series data (being not the case for the often used Cholesky approach). It would be interesting to study whether inference could be developed for other matrix-valued objects, such as curves of positive semi-definite matrices.

Acknowledgements. We thank the Associate Editor for all encouraging comments that helped to improve our paper. Further thanks go to Gabriel Bailly for having found a slight but important mistake in the first version of our code for the numerical simulation studies. Johannes Krebs gratefully acknowledges the support of the Deutsche Forschungsgemeinschaft (grants KR-4977/1-1, KR-4977/2-1) and the hospitality of ISBA/LIDAM (UCLouvain).

References

A. Antoniadis, G. Grégoire, and I. McKeague. Wavelet methods for curve estimation. *J. Amer. Statist. Assoc.*, 89:1340–1353, 1994.

M. Arnaudson, F. Barbaresco, and L. Yang. Riemannian medians and means with applications to radar signal processing. *IEEE J. Sel. Top. Signal Proc.*, 7:595–604, 2013.

V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.*, 29(1):328–347, 2007.

N. Boumal and P.-A. Absil. Discrete regression methods on the cone of positive-definite matrices. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4232–4235. IEEE, 2011.

R. Caseiro, J. Henriques, P. Martins, and J. Batista. A nonparametric Riemannian framework on tensor field with application to foreground segmentation. *Pattern Recognit.*, 45:3997–4017, 2012.

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

J. Chau and R. von Sachs. Functional mixed effects wavelet estimation for spectra of replicated time series. *Electron. J. Stat.*, 10(2):2461–2510, 2016.

J. Chau and R. von Sachs. Intrinsic wavelet regression for curves of Hermitian positive definite matrices. *J. Amer. Statist. Assoc.*, 116(534):819–832, 2021.

J. Chau and R. von Sachs. Time-varying spectral matrix estimation via intrinsic wavelet regression for surfaces of Hermitian positive definite matrices. *Computational Statistics and Data Analysis*, 174:107477, 2022.

I. Daubechies and J. C. Lagarias. Two-scale difference equations. i. existence and global regularity of solutions. *SIAM J. Math. Anal.*, 22(5):1388–1410, 1991.

I. Daubechies and J. C. Lagarias. Sets of matrices all infinite products of which converge. *Linear Algebra Appl.*, 161(C):227–263, 1992a.

I. Daubechies and J. C. Lagarias. Two-scale difference equations ii. local regularity, infinite products of matrices and fractals. *SIAM J. Math. Anal.*, 23 (4):1031–1079, 1992b.

D. L. Donoho. Smooth wavelet decompositions with blocky coefficient kernels. In *Recent Advances in Wavelet Analysis*, pages 1–43. Academic Press, 1993.

I. L. Dryden, A. Koloydenko, and D. Zhou. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.*, 3(3):1102–1123, 2009.

P. Fillard, V. Arsigny, X. Pennec, K. Hayasghi, P. Thompson, and N. Ayache. Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines. *Neuroimage*, 34:639–650, 2007.

P. Fletcher and S. Joshib. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Process.*, 87:250–262, 2007.

K. Friston. Functional and effective connectivity: a review. *Brain Connect.*, 1: 13–36, 2011.

P. Hall. On bootstrap confidence intervals in nonparametric regression. *Ann. Statist.*, 20(2):695–711, 1992.

W. Härdle and A. W. Bowman. Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.*, 83 (401):102–110, 1988.

J. Hinkle, P. T. Fletcher, and S. Joshi. Intrinsic polynomials for regression on Riemannian manifolds. *J. Math. Imaging Vis.*, 50(1-2):32–52, 2014.

M. H. Jansen and P. J. Oonincx. *Second Generation Wavelets and Applications*. Springer, 2005.

R. Klees and R. Haagmans. *Wavelets in the Geosciences*, volume 90. Springer, 2000.

J. M. Lee. *Riemannian Manifolds - An Introduction to Curvature*. Graduate Texts in Mathematics, 176. Springer, New York, 1997.

Z. Lin. Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIAM J. Matrix Anal. Appl.*, 40(4):1353–1370, 2019.

X. Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *J. Math. Imaging Vis.*, 25(1):127, 2006.

I. U. Rahman, I. Drori, V. C. Stodden, D. L. Donoho, and P. Schröder. Multiscale representations for manifold-valued data. *Multiscale Model. Simul.*, 4 (4):1201–1232, 2005.

L. W. Tu. *Differential Geometry*. Graduate Texts in Mathematics, 275. Springer, New York, 2017.

Y. Yuan, H. Zhu, W. Lin, and J. Marron. Local polynomial regression for symmetric positive definite matrices. *J. R. Stat. Soc., Ser. B*, 74(4):697–719, 2012.

H. Zhu, Y. Chen, J. G. Ibrahim, Y. Li, C. Hall, and W. Lin. Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *J. Amer. Statist. Assoc.*, 104(487):1203–1212, 2009.

Appendix A: Technical details on Section 2

A.1. Fréchet means

Let $\mathbb{E}[X]$ denote the Fréchet (or Karcher) mean of a random element X on $Sym^+(d)$, which provides a general notion of location in metric spaces and is in our case defined as the set

$$\mathbb{E}[X] := \underset{R\in Sym^+(d)}{\arg\min}\ \mathbb{E}[d^2(R,X)] = \underset{R\in Sym^+(d)}{\arg\min} \int_{Sym^+(d)} d^2(R,S)\,\mathrm{d}\mathbb{P}^X(S).$$

If $\mathbb{P}^X \in \mathcal{P}_2(Sym^+(d))$, then at least one such point exists. Moreover, the Fréchet mean $\mathbb{E}[X]$ is unique because $Sym^+(d)$ is a geodesically complete and flat manifold with respect to the log-Euclidean metric. Since at every point $S \in Sym^+(d)$ the exponential map (see Table 7) is defined on the entire tangent space $T_S Sym^+(d)$, the Fréchet mean $\mathbb{E}[X] \in Sym^+(d)$ is also uniquely characterized by the condition $\mathbb{E}[\mathrm{Log}_{\mathbb{E}[X]}(X)] = 0$. The simple geometry under the log-Euclidean metric then allows to conclude

$$\mathbb{E}[X] = \exp\left(\mathbb{E}[\log(X)]\right) \ , \tag{A.1}$$

where $\mathbb{E}[\log(X)]$ is just the usual entrywise Euclidean expectation of the random matrix $\log(X)$ in the Hilbert space space $(Sym(d), \langle\cdot,\cdot\rangle)$, i.e. the unique element such that $\mathbb{E}\langle \log X, A\rangle = \langle \mathbb{E}[\log X], A\rangle$ for all $A \in Sym(d)$, which yields $(\mathbb{E}[\log X])_{ij} = \mathbb{E}[(\log X)_{ij}]$. If X has a discrete distribution $\mathbb{P}^X = \sum_{i=1}^n w_i \delta_{S_i}$, where $\sum_i w_i = 1$, $w_i > 0$ and $S_i \in Sym^+(d)$, then

$$\mathbb{E}[X] = \exp\left(\sum_{i=1}^n w_i \log(S_i)\right) =: \mathrm{Ave}(\{S_i\}; \{w_i\}), \tag{A.2}$$

see also Arsigny et al. [2007]. In case $w_i = 1/n$ we may simply write $\mathrm{Ave}(\{S_i\})$ (or $\overline{S}_n$ resp. $\overline{X}_n$) for the average $\mathrm{Ave}(\{S_i\}; \{1/n\})$.

Furthermore, if we consider a smooth curve $c\colon \mathbb{R} \supset I \to Sym^+(d)$ as a mapping from the probability space $(I, \mathcal{B}(I), \lambda/\lambda(I))$ to the 1-dimensional submanifold $c(I) \subset Sym^+(d)$ we can also define the *intrinsic mean* $\mathrm{Ave}_I(c)$ of c over the interval I by

$$\mathrm{Ave}_I(c) := \operatorname*{arg\,min}_{R \in c(I)} \int_{c(I)} d^2(R, S)\, \mathrm{d}\nu^c(S)$$

with $\nu^c := \lambda/\lambda(I) \circ c^{-1}$. Again, in case $\nu^c \in \mathcal{P}_2(c(I))$, the intrinsic mean $\mathrm{Ave}_I(c)$ of c exists and is also unique. Restricting the above considerations on the submanifold $c(I)$ we then obtain

$$\mathrm{Ave}_I(c) = \exp\left\{\int_{c(I)} \log(S)\, \mathrm{d}\nu^c(S)\right\} = \exp\left\{\frac{1}{\lambda(I)} \int_I \log(c(t))\, \mathrm{d}t\right\}. \tag{A.3}$$

A.2. Illustration of midpoint pyramid in Section 2.3

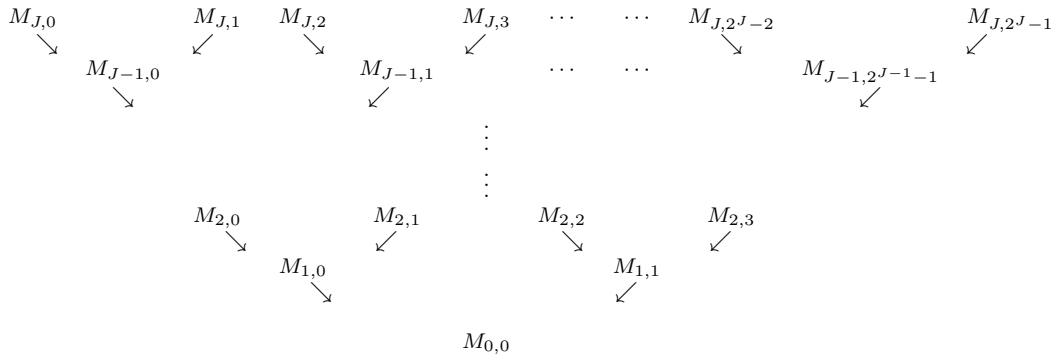


FIG 6. *Scheme of the midpoint pyramid algorithm. Each $M_{j,k}$ is obtained from $M_{j+1,2k}$ and $M_{j+1,2k+1}$ by the averaging procedure in (2.6), resulting in a consecutive finer average for a fixed $M_{j,k}$ in the case that the finest scale J increases to ∞.*

A.3. Convergence of AI refinement schemes

Proof of Lemma 2.2. Let $m = 1$, then $p = d_{j+1}(p) \in \{0, 1\}$ and (2.11) respectively (2.12) yield

$$\begin{bmatrix} \log \widetilde{M}_{j+1,2k+p-2L} \\ \vdots \\ \log \widetilde{M}_{j+1,2k+p+2L} \end{bmatrix} = (E_N \mathbb{1}\{d_{j+1}(p) = 0\} + O_N \mathbb{1}\{d_{j+1}(p) = 1\})_{\diagdown_d} \begin{bmatrix} \log M_{j,k-2L} \\ \vdots \\ \log M_{j,k+2L} \end{bmatrix}.$$

We establish the general case via induction ($m \rightsquigarrow m+1$): Let $p \in \{0, \dots, 2^{m+1} - 1\}$ with unique dyadic expansion

$$p = \sum_{i=j+1}^{j+m+1} d_i(p) 2^{j+m+1-i} = 2 \sum_{i=j+1}^{j+m} d_i(p) 2^{j+m-i} + d_{j+m+1}(p).$$

Observe that

$$q = \sum_{i=j+1}^{j+m} d_i(p) 2^{j+m-i} \in \{0, \dots, 2^m - 1\},$$

so by (2.11) respectively (2.12) and the induction hypothesis we can conclude

$$\begin{bmatrix} \log \widetilde{M}_{j+m+1,2^{m+1}k+p-2L} \\ \vdots \\ \log \widetilde{M}_{j+m+1,2^{m+1}k+p+2L} \end{bmatrix} = \begin{bmatrix} \log \widetilde{M}_{j+m+1,2(2^m k+q)+d_{j+m+1}(p)-2L} \\ \vdots \\ \log \widetilde{M}_{j+m+1,2(2^m k+q)+d_{j+m+1}(p)+2L} \end{bmatrix} =$$

$$= (E_N \mathbb{1}\{d_{j+m+1}(p) = 0\} + O_N \mathbb{1}\{d_{j+m+1}(p) = 1\})_{\diagdown d} \begin{bmatrix} \log \widetilde{M}_{j+m,2^m k+q-2L} \\ \vdots \\ \log \widetilde{M}_{j+m,2^m k+q+2L} \end{bmatrix} =$$

$$= (U_{m+1} U_m \cdot \ldots \cdot U_1)_{\diagdown d} \begin{bmatrix} \log M_{j,k-2L} \\ \vdots \\ \log M_{j+m,k+2L} \end{bmatrix}.$$

□

Below we give the explicit forms of the matrices E_N and O_N as introduced in Section 2.4:

$$E_N := \begin{bmatrix}
-c_L & -c_{L-1} & \cdots & -c_1 & 1 & c_1 & \cdots & c_{L-1} & c_L & 0 & & & & & \\
c_L & c_{L-1} & \cdots & c_1 & 1 & -c_1 & \cdots & -c_{L-1} & -c_L & 0 & & & & & \\
0 & -c_L & \cdots & -c_2 & -c_1 & 1 & \cdots & c_{L-2} & c_{L-1} & c_L & & & & & \\
0 & c_L & \cdots & c_2 & c_1 & 1 & \cdots & -c_{L-2} & -c_{L-1} & -c_L & & & & & \\
 & & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & & & \\
 & & & 0 & -c_L & -c_{L-1} & \cdots & -c_1 & 1 & c_1 & \cdots & c_{L-1} & c_L & 0 & \\
 & & & 0 & c_L & c_{L-1} & \cdots & c_1 & 1 & -c_1 & \cdots & -c_{L-1} & -c_L & 0 & \\
 & & & & & & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \\
 & & & & & & 0 & -c_L & -c_{L-1} & -c_{L-2} & \cdots & 1 & c_1 & \cdots & c_L & 0 \\
 & & & & & & 0 & c_L & c_{L-1} & c_{L-2} & \cdots & 1 & -c_1 & \cdots & -c_L & 0 \\
 & & & & & & & 0 & -c_L & -c_{L-1} & \cdots & -c_1 & 1 & \cdots & c_{L-1} & c_L
\end{bmatrix} \tag{E_N}$$

$$O_N := \begin{bmatrix}
c_L & c_{L-1} & \cdots & c_1 & 1 & -c_1 & \cdots & -c_{L-1} & -c_L & 0 & & & & & & \\
0 & -c_L & \cdots & -c_2 & -c_1 & 1 & \cdots & c_{L-2} & c_{L-1} & c_L & & & & & & \\
0 & c_L & \cdots & c_2 & c_1 & 1 & \cdots & -c_{L-2} & -c_{L-1} & -c_L & & & & & & \\
 & & & & \vdots & \vdots & & \vdots & \vdots & \vdots & & & & & & \\
 & & & 0 & c_L & c_{L-1} & \cdots & c_1 & 1 & -c_1 & \cdots & -c_{L-1} & -c_L & 0 & & \\
 & & & 0 & 0 & -c_L & \cdots & -c_2 & -c_1 & 1 & \cdots & c_{L-2} & c_{L-1} & c_L & & \\
 & & & & & & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \\
 & & & & & & & -c_L & -c_{L-1} & -c_{L-2} & \cdots & c_1 & c_2 & c_3 & \cdots & c_L & 0 \\
 & & & & & & & c_L & c_{L-1} & c_{L-2} & \cdots & -c_1 & -c_2 & -c_3 & \cdots & -c_L & 0 \\
 & & & & & & & 0 & -c_L & -c_{L-1} & \cdots & 1 & c_1 & c_2 & \cdots & c_{L-1} & c_L \\
 & & & & & & & 0 & c_L & c_{L-1} & \cdots & 1 & -c_1 & -c_2 & \cdots & -c_{L-1} & -c_L
\end{bmatrix} \tag{O_N}$$

Appendix B: Technical details on wavelet regression in Section 3

Proposition B.1 (Rate of convergence from Chau and von Sachs [2021]). *Consider the wavelet estimator $\widehat{M}_{J,k,n}(J_0)$, based on a refinement scheme of order N and on (linear) thresholding on scale J_0. Let further $c(t)$ be a smooth curve of SPD matrices, possessing at least N continuous derivatives. Then there is a $C \in \mathbb{R}_+$, which does not depend on n and k such that*

$$\mathbb{E}[\| \log \widehat{M}_{J,k,n}(J_0) - \mathbb{E}[\log \widehat{M}_{J,k,n}(J_0)]\|_F^2] \leq C 2^{-(J-J_0)}$$

and

$$\|\mathbb{E}[\log \widehat{M}_{J,k,n}(J_0)] - \log M_{J,k}\|_F \leq C 2^{-NJ_0}.$$

In particular, the choice $J_0 = \lfloor J/(2N+1) \rfloor$ balances variance and squared bias, so

$$\mathbb{E}[d(\widehat{M}_{J,k,n}(J_0), M_{J,k})^2] = \mathbb{E}[\| \log \widehat{M}_{J,k,n}(J_0) - \log M_{J,k}\|_F^2] \leq 2Cn^{-2N/(2N+1)} \ ,$$

which implies that $n^{-1} \sum_{k=0}^{n-1} \mathbb{E}[d(\widehat{M}_{J,k,n}(J_0), M_{J,k})^2] = O(n^{-2N/(2N+1)})$.

Lemma B.2. *Let $t = k2^{-J}$ be a dyadic rational in $[0,1]$, then*

$$\|\mathcal{C}(t)\|_{\mathcal{L}} \leq \mathbb{E}\left[\| \log M_{J,k,n}\|_F^2\right]$$

and, in particular,

$$\sup_{t\in[0,1]} \|\mathcal{C}(t)\|_{\mathcal{L}} \leq \sup_{\substack{k=0,\dots,2^J-1,\\ J\geq 0}} \mathbb{E}\left[\| \log M_{J,k,n}\|_F^2\right].$$

Proof. Let $A \in Sym(d)$ with $\|A\|_F \leq 1$ and let $(E_i)_i$ be an orthonormal basis of the Hilbert space $(Sym(d), \langle \cdot, \cdot \rangle_F)$. Then, for a dyadic rational $t = k2^{-J} \in [0,1]$,

$$\begin{aligned}
\|\mathcal{C}(t)A\|_F &= \|\mathbb{E}\langle A, \log M_{J,k,n} - \mathbb{E}[\log M_{J,k,n}]\rangle(\log M_{J,k,n} - \mathbb{E}[\log M_{J,k,n}])\|_F \\
&\leq \mathbb{E}\left[\| \log M_{J,k,n} - \mathbb{E}[\log M_{J,k,n}]\|_F^2\right] \\
&= \operatorname{trace}(\mathcal{C}(t)) \\
&= \sum_i \langle \mathcal{C}(t)E_i, E_i\rangle_F \\
&= \mathbb{E}\sum_i \langle \log M_{J,k,n}, E_i\rangle_F^2 \\
&= \mathbb{E}\left[\| \log M_{J,k,n}\|_F^2\right]
\end{aligned}$$

and taking the supremum yields $\|\mathcal{C}(t)\|_{\mathcal{L}} \leq \mathbb{E}\| \log M_{J,k,n}\|_F^2$. The dyadic numbers are of course dense in $[0,1]$ and therefore the amendment follows from the càdlàg property of $\mathcal{C}$. □

In order to establish asymptotic normality, we rely on Lyapunov's condition.

Theorem B.3 (Lyapunov condition)**.** *Let $(\xi_{n,i})_{i=1}^{\ell_n}$ be a sequence of independent real-valued random variables for each $n \in \mathbb{N}$ such that $\mathbb{E}[\xi_{n,i}] = 0$ and $\mathbb{E}[|\xi_{n,i}|^2] < \infty$. If the partial sums $S_n = \sum_{i=1}^{\ell_n} \xi_{n,i}$ satisfy, for some $\delta > 0$,*

$$\lim_{n\to\infty} \frac{1}{\mathrm{Var}(S_n)^{1+\delta/2}} \sum_{i=1}^{\ell_n} \mathbb{E}\left[|\xi_{n,i}|^{2+\delta}\right] = 0, \tag{B.1}$$

then $S_n/\sqrt{\mathrm{Var}(S_n)} \Rightarrow \mathcal{N}(0,1)$ as $n \to \infty$. In particular, if $\mathrm{Var}(S_n) \to \sigma^2$, then $S_n \Rightarrow \mathcal{N}(0,\sigma^2)$.

Proof of Proposition 3.5. The proof is divided in two parts, the first part covering the statement for the fixed dyadic number $k2^{-j}$, the second the statement for $k_n 2^{-J_{0,n}} \to x$.

(a) We rely on the Cramér-Wold-device, i.e. we show for arbitrary $A \in Sym(d)$, that

$$2^{-j/2} n^{1/2} \langle \log M_{j,k,n} - \mathbb{E}[\log M_{j,k,n}], A\rangle \quad \Rightarrow \quad \mathcal{N}(0, \langle \Sigma_{j,k} A, A\rangle).$$

To that end, first observe that inserting representation (3.10) shows

$$\begin{aligned} &2^{-j/2} n^{1/2} \langle \log M_{j,k,n} - \mathbb{E}[\log M_{j,k,n}], A\rangle = \\ &= 2^{-(J-j)/2} \sum_{u=0}^{2^{J-j}-1} \left\langle \log M_{J,u+k2^{J-j},n} - \mathbb{E}[\log M_{J,u+k2^{J-j},n}], A\right\rangle. \end{aligned}$$

Thus, it suffices to verify the conditions stated in Theorem B.3 for the random variables

$$\xi_{n,u} := 2^{-(J-j)/2} \left\langle \log M_{J,u+k2^{J-j},n} - \mathbb{E}[\log M_{J,u+k2^{J-j},n}], A\right\rangle,$$

where $0 \le u \le 2^{J-j} - 1 =: \ell_n - 1$. To that end, let $S_n := \sum_{u=0}^{\ell_n - 1} \xi_{n,u}$. From (3.11) and (3.13) we obtain directly

$$\begin{aligned} \mathrm{Var}\left(S_n\right) = \sum_{u=0}^{\ell_n-1} \mathrm{Var}(\xi_{n,u}) = 2^{-(J-j)} \sum_{u=0}^{2^{J-j}-1} \left\langle \mathcal{C}(u2^{-J} + k2^{-j})A, A\right\rangle \quad \xrightarrow{n\to\infty} \\ \left\langle 2^j \int_{k2^{-j}}^{(k+1)2^{-j}} \mathcal{C}(t)\ \mathrm{d}tA, A \right\rangle = \langle \Sigma_{j,k} A, A\rangle. \end{aligned}$$

Furthermore, for $\delta = r - 2 > 0$, the uniform bounded moments condition 3.1 yields

$$\begin{aligned} \sum_{u=0}^{\ell_n-1} \mathbb{E}\left[|\xi_{n,u}|^{2+\delta}\right] &= \sum_{u=0}^{\ell_n-1} \ell_n^{-(1+\delta/2)} \mathbb{E}\left[\left|\langle \log M_{J,u+k2^{J-j},n} - \mathbb{E}[\log M_{J,u+k2^{J-j},n}], A\rangle\right|^{2+\delta}\right] \\ &\le \sum_{u=0}^{\ell_n-1} \ell_n^{-(1+\delta/2)}\ \|A\|^r\ \mathbb{E}\left[\left\|\log M_{J,u+k2^{J-j},n} - \mathbb{E}[\log M_{J,u+k2^{J-j},n}]\right\|^r\right] \end{aligned}$$

$$\leq \ell_n^{-(1+\delta/2)} \, \|A\|^r \sum_{u=0}^{\ell_n-1} 2^r \, \mathbb{E}\left[\|\log M_{J,u+k2^{J-j},n}\|^r\right] \leq 2^r \, \|A\|^r \, K(r) \, \ell_n^{-\delta/2}.$$

Because $\langle \Sigma_{j,k} A, A\rangle > 0$, this establishes the Lyapunov condition (B.1):

$$\frac{1}{\mathrm{Var}(S_n)^{1+\delta/2}} \sum_{u=0}^{\ell_n-1} \mathbb{E}\left[|\xi_{n,u}|^{2+\delta}\right] \leq \frac{2^r \, \|A\|^r \, K(r)}{\mathrm{Var}(S_n)^{1+\delta/2}} \, 2^{-(J-j)\delta/2} \xrightarrow{n\to\infty} 0.$$

(b) To keep the notation simple, we omit the dependence of J_0 and k on n in the following. We can proceed similar to the proof of (a) by substituting J_0 for j. That is, it suffices to verify the conditions of Theorem B.3 for the random variables

$$\xi_{n,u} := 2^{-(J-J_0)/2} \left\langle \log M_{J,u+k2^{J-J_0},n} - \mathbb{E}[\log M_{J,u+k2^{J-J_0},n}], A \right\rangle,$$

where $0 \leq u \leq 2^{J-J_0} - 1 =: \ell_n - 1$ and $A \in Sym(d)$ is arbitrary. To that end, let $S_n = \sum_{u=0}^{\ell_n-1} \xi_{n,u}$. Then, using (3.11) again, we have

$$\mathrm{Var}(S_n) = \sum_{u=0}^{\ell_n-1} \mathrm{Var}(\xi_{n,u}) \; =$$

$$= \langle \mathcal{C}(x)A, A\rangle + 2^{-(J-J_0)} \sum_{u=0}^{2^{J-J_0}-1} \left\langle \left(\mathcal{C}(u2^{-J} + k2^{-J_0}) - \mathcal{C}(x)\right)A, A\right\rangle = \langle \mathcal{C}(x)A, A\rangle + R(n).$$

Now let $\varepsilon > 0$ be arbitrary. Then by continuity of $\mathcal{C}$ in x, there exists a $\delta > 0$, such that $\|\mathcal{C}(x) - \mathcal{C}(y)\|_{\mathcal{L}} < \varepsilon$ for all $y \in (x - \delta, x + \delta)$. Since

$$|u2^{-J} + k2^{-J_0} - x| \leq |k2^{-J_0} - x| + u2^{-J} \leq |k2^{-J_0} - x| + 2^{-J_0} \xrightarrow{n\to\infty} 0,$$

we can assume n large enough, such that $|u2^{-J} + k2^{-J_0} - x| < \delta$ for all $u = 0, \ldots, \ell_n - 1$. Therefore

$$|R(n)| \leq \|A\|^2 2^{-(J-J_0)} \sum_{u=0}^{2^{J-J_0}-1} \|\mathcal{C}(u2^{-J} - k2^{-J_0}) - \mathcal{C}(x)\|_{\mathcal{L}} < \|A\|^2 \varepsilon,$$

which implies $|R(n)| \to 0$ and hence $\mathrm{Var}(S_n) \to \langle \mathcal{C}(x)A, A\rangle > 0$. In order to establish the Lyapunov condition (B.1), let $\delta = r - 2 > 0$. Just like in the proof of (a), we then have

$$\sum_{u=0}^{\ell_n-1} \mathbb{E}\left[|\xi_{n,u}|^{2+\delta}\right] \leq 2^r \, \|A\|^r \, K(r) \, \ell_n^{-\delta/2},$$

which in turn immediately yields Lyapunov's condition in the same way as in (a). □

Proving Theorem 3.6 will require to study the AI wavelet estimator at a specific point $x \in [0,1]$. To that end, let us first introduce some terminology regarding dyadic approximations in the unit interval. Let $x \in [0,1]$ have the dyadic representation $(x)_2 = (d_0, d_1, \ldots)$ with $d_i \in \{0,1\}$, that is

$$x = \sum_{i=0}^{\infty} d_i 2^{-i}.$$

In order to guaranty uniqueness of the dyadic representation we agree to choose the the shortened representation $(x)_2 = (d_0, d_1, \ldots, d_{j-1}, 1, 0, 0, \ldots)$ for dyadic rationals $x = \sum_{i=0}^{j-1} d_i 2^{-i} + 2^{-j}$. (Note, that we could also use the representation $(x)_2 = (d_0, d_1, \ldots, d_j, 0, 1, 1, \ldots)$ instead; this does not change the arguments much.) Now we define, for $x \in [0,1]$ with dyadic representation $(x)_2 = (d_0, d_1, \ldots)$ and fixed scale $j \in \mathbb{N}$, the dyadic approximation of x via $(x)_2^j = (d_0, d_1, \ldots, d_j, 0, 0, \ldots)$. This means, $(x)_2^j$ the dyadic representation of the rational approximation

$$\widehat{x} = \sum_{i=0}^{j} d_i 2^{-i} = d_0 2^0 + d_1 2^{-1} + \ldots + d_j 2^{-j} = \left(d_0 2^j + d_1 2^{j-1} + \ldots + d_j 2^0\right)\ 2^{-j} =: k^{(j)}\ 2^{-j}.$$

Note that $k^{(j)} \in \{0, \ldots, 2^j - 1\}$ is ensured, since $d_0 = 1$ implies $d_j = 0$ for $j \geq 1$. Furthermore, for a smaller scale $j_0 \leq j$ we have

$$k^{(j)} = \sum_{i=0}^{j} d_i 2^{j-i} = d_j 2^0 + d_{j-1} 2^1 + \ldots + d_{j_0+1} 2^{j-j_0-1} + \left(\sum_{i=0}^{j_0} d_{j_0-i} 2^i\right) 2^{j-j_0}, \tag{B.2}$$

so we can also express the rational approximation as

$$\widehat{x} = k^{(j)} 2^{-j} = d_j 2^{-j} + d_{j-1} 2^{-(j-1)} + \ldots + d_{j_0+1} 2^{-(j_0+1)} + k_{j,j_0} 2^{-j_0} \quad \text{with} \quad k_{j,j_0} = \sum_{i=0}^{j_0} d_{j_0-i} 2^i. \tag{B.3}$$

Proof of Theorem 3.6. To keep the notation simple, we omit the dependence of J_0 and k on n. Let $x \in (0,1)$ with dyadic representation $(x)_2 = (d_0(x), d_1(x), \ldots)$ and consider the rational approximation (B.3) with respect to J and J_0,

$$\widehat{x} = k^{(J)} 2^{-J} = d_J(x) 2^{-J} + \ldots + d_{J_0+1} 2^{-(J_0+1)} + k,$$

where $k^{(J)} = \sum_{i=0}^{J} d_i(x) 2^{J-i}$ and

$$k = k_{J,J_0} = \sum_{i=0}^{J_0} d_{J_0-i}(x) 2^i \in \{0, \ldots, 2^{J_0} - 1\}.$$

Let $\ell = J - J_0$ and observe

$$2^\ell k = 2^{J-J_0} \sum_{i=0}^{J_0} d_{J_0-i}(x) 2^i = \sum_{i=0}^{J_0} d_i(x) 2^{J-i} = k^{(J)} - \sum_{i=J_0+1}^{J} d_i(x) 2^{J-i} =: k^{(J)} - p.$$

By construction $p \in \{0, \dots, 2^\ell - 1\}$, so an application of Lemma 2.2 yields

$$\begin{bmatrix} \log \widehat{M}_{J,k^{(J)}-2L,n}(J_0) \\ \vdots \\ \log \widehat{M}_{J,k^{(J)}+2L,n}(J_0) \end{bmatrix} = \begin{bmatrix} \log \widehat{M}_{J,2^\ell k+p-2L,n}(J_0) \\ \vdots \\ \log \widehat{M}_{J,2^\ell k+p+2L,n}(J_0) \end{bmatrix} =$$

$$= \left(U_\ell(x)U_{\ell-1}(x) \cdot \ldots \cdot U_1(x)\right)_{\searrow d} \begin{bmatrix} \log M_{J_0,k-2L,n} \\ \vdots \\ \log M_{J_0,k+2L,n} \end{bmatrix} \tag{B.4}$$

with

$$U_i(x) = E_N \mathbb{1}\left\{d_{J_0+i}(x) = 0\right\} + O_N \mathbb{1}\left\{d_{J_0+i}(x) = 1\right\}, \quad i = 1, \dots, \ell.$$

In particular, for $i \in \{-2L, \dots, 2L\}$, it follows from (B.4) and the representation (3.10) that

$$\log \widehat{M}_{J,k^{(J)}+i,n}(J_0) = \sum_{v=-2L}^{2L} \alpha_{i,v} \log M_{J_0,k+v,n} = \sum_{v=-2L}^{2L} \alpha_{i,v}\, 2^{-\ell} \sum_{u=0}^{2^\ell-1} \log M_{J,u+(k+v)2^\ell,n}, \tag{B.5}$$

where of course $(\alpha_{i,-2L}, \dots, \alpha_{i,2L})$ is the $(2L+i+1)$th row of $U_\ell(x) \cdots U_1(x)$. Naturally the coefficients $(\alpha_{i,v})_v$ depend on J_0 and J (resp. n), but we omit this dependence in the notation in the following. Since the rows of the matrices E_N and O_N sum up to 1 (see (E_N) and (O_N)) it is elementary to see that this still holds for arbitrary products of these matrices and hence $\sum_v \alpha_{i,v} = 1$ as well. Note that this in turn implies immediately

$$\sum_{v=-2L}^{2L} \alpha_{i,v}^2 \geq 1/(4L+1)$$

uniformly in n, which follows from the minimization problem

$$\min_{w_1,\dots,w_L,\lambda} \left\{ \sum_v w_v^2 + \lambda \left(\sum_v w_v - 1 \right) \right\}.$$

Also, since $x \in (0,1)$ is assumed to be a dyadic rational, i.e. $x = q2^{-j}$ for some $j \geq 0$ and $q \in \{1, \dots, 2^j - 1\}$, the lifting in B.4 will be carried out ultimately only with

$$U_1(x) = \ldots = U_\ell(x) = E_N$$

once J and J_0 have exceeded the scale j, because then $d_{J_0+1}(x) = \ldots = d_J(x) = 0$. Therefore, for dyadic rationals $x \in (0,1)$, the limit

$$\lim_{n\to\infty} \sum_{v=-2L}^{2L} \alpha_{i,v}^2 = \sum_{v=-2L}^{2L} \lim_{n\to\infty} (U_\ell(x)\cdots U_1(x))_{2L+i+1,\nu}^2 = \sum_{v=-2L}^{2L} (E_N^\infty)_{2L+i+1,\nu}^2 = \sum_{v=-2L}^{2L} \varphi_L^2(v) \tag{B.6}$$

exists due to the considerations in section 2.4, in particular the discussion leading to (2.15).

Now, in order to establish the assertion, we can use the representation (B.5) and rely on the Cramér-Wold-device in conjunction with Theorem B.3 once again, i.e. we show, for arbitrary $A \in Sym(d)$, that

$$\begin{aligned}
&\left\langle 2^{J_0/2} n^{1/2} \left(\log \widehat{M}_{J,k^{(J)},n}(J_0) - \mathbb{E}\left[\log \widehat{M}_{J,k^{(J)},n}(J_0) \right] \right),\ A \right\rangle \\
&= \sum_{u=0}^{2^\ell-1} \sum_{v=-2L}^{2L} 2^{-\ell/2}\ \alpha_{0,v} \left\langle \log M_{J,u+(k+v)2^\ell,n} - \mathbb{E}\left[\log M_{J,u+(k+v)2^\ell,n} \right],\ A \right\rangle
\end{aligned}$$

$$\Rightarrow \quad \mathcal{N}(0, \langle \kappa_N \mathcal{C}(x) A, A \rangle). \tag{B.7}$$

Taking a closer look shows that the indices $u + (k+v)2^\ell$ stemming from the double sum are all pairwise different, i.e. for $u_1 \neq u_2 \in \{0, \ldots, 2^\ell - 1\}$,

$$\left\{ u_1 + (k+v)2^\ell \;\middle|\; v = -2L, \ldots, 2L \right\} \cap \left\{ u_2 + (k+v)2^\ell \;\middle|\; v = -2L, \ldots, 2L \right\} = \emptyset.$$

Therefore, the inner sums in (B.7),

$$\xi_{n,u} := 2^{-\ell/2} \sum_{v=-2L}^{2L} \alpha_{0,v} \left\langle \log M_{J,u+(k+v)2^\ell,n} - \mathbb{E}\left[\log M_{J,u+(k+v)2^\ell,n} \right],\ A \right\rangle, \quad u = 0, \ldots, 2^\ell - 1 \tag{B.8}$$

are actually pairwise independent with $\mathbb{E}[\xi_{n,u}] = 0$ and, due to the covariance identity(3.7),

$$\begin{aligned}
\mathrm{Var}(\xi_{n,u}) &= 2^{-\ell} \sum_{v=-2L}^{2L} \alpha_{0,v}^2\ \mathrm{Var}\left(\left\langle \log M_{J,u+(k+v)2^\ell,n} - \mathbb{E}\left[\log M_{J,u+(k+v)2^\ell,n} \right],\ A \right\rangle \right) \\
&= 2^{-\ell} \sum_{v=-2L}^{2L} \alpha_{0,v}^2 \left\langle \mathcal{C}\left(u2^{-J} + (k+\nu)2^{-J_0} \right) A,\ A \right\rangle.
\end{aligned}$$

Let $S_n := \sum_{u=0}^{2^\ell-1} \xi_{n,u}$, then by independence

$$\mathrm{Var}\left(S_n\right) = \sum_{u=0}^{2^\ell-1} 2^{-\ell} \sum_{v=-2L}^{2L} \alpha_{0,v}\ \left\langle \mathcal{C}\left(u2^{-J} + (k+v)2^{-J_0} \right) A,\ A \right\rangle$$

and, since

$$u2^{-J} + (k+v)2^{-J_0} = u2^{-J} + \left(\sum_{i=0}^{J_0} d_{J_0-i}(x)2^i + v\right) 2^{-J_0}$$
$$= u2^{-J} + \sum_{i=0}^{J_0} d_i(x)2^{-i} + v2^{-J_0} \quad \xrightarrow{n\to\infty} \quad x,$$

continuity of $\mathcal{C}$ in x yields, by a similar argument as in the proof of Proposition 3.5 (b), that

$$\mathrm{Var}\,(S_n) \quad \xrightarrow{n\to\infty} \quad \sum_{v=-2L}^{2L} \varphi_L^2(v)\ \langle \mathcal{C}(x)\ A,\ A\rangle = \langle \kappa_N\ \mathcal{C}(x)\ A,\ A\rangle\,. \tag{B.9}$$

Moreover, for $\delta = r - 2 > 0$, the uniform bounded moments condition (3.1) yields

$$\sum_{u=0}^{2^\ell-1} \mathbb{E}[|\xi_{n,u}|^{2+\delta}]$$
$$\le \sum_{u=0}^{2^\ell-1} (2^\ell)^{-(1+\delta/2)} \sum_{v=-2L}^{2L} |\alpha_{0,v}|^{2+\delta} \mathbb{E}\left\|\log M_{J,u+(k+v)2^\ell,n} - \mathbb{E}[\log M_{J,u+(k+v)2^\ell,n}]\right\|^{2+\delta} \|A\|^{2+\delta}$$
$$\le (2^\ell)^{-\delta/2}\ 2^r\ \|A\|^r\ K(r)\ \max_v(|\alpha_{0,v}|^2)^{\delta/2} \sum_{v=-2L}^{2L} |\alpha_{0,v}|^2$$
$$\le (2^\ell)^{-\delta/2}\ 2^r\ \|A\|^r\ K(r)\ \left(\sum_{v=-2L}^{2L} |\alpha_{0,v}|^2\right)^{1+\delta/2} \quad \to \quad 0,$$

which in turn verifies the Lyapunov condition (B.1), due to (B.9) and $\langle \kappa_N\ \mathcal{C}(x)\ A,\ A\rangle > 0$. Hence, the conditions of Theorem B.3 are satisfied for $(\xi_{n,u})$ and (B.7) is proven. □

Proof of Proposition 3.8. The result follows in a similar spirit as Theorem 3.6. Let $\delta > 0$ such that the uniform bounded moments condition in (3.1) is satisfied for $2+\delta$. The Lyapunov condition for the normalized estimator

$$\mathrm{Var}(\langle A, \log \widehat{M}_{J,k,n}(J_0) - \mathbb{E}[\log \widehat{M}_{J,k,n}(J_0)]\rangle^{-1/2} \cdot \langle A, \log \widehat{M}_{J,k,n}(J_0) - \mathbb{E}[\log \widehat{M}_{J,k,n}(J_0)]\rangle$$

can be established with the representation from (B.5) and reads as follows

$$\frac{\sum_{v=-2L}^{2L}(|\alpha_v|\ 2^{-\ell})^{2+\delta} \sum_{u=0}^{2^\ell-1} \mathbb{E}[|\langle A, \log M_{J,u+(k_0+v)2^\ell,n} - \mathbb{E}[\log M_{J,u+(k_0+v)2^\ell,n}]\rangle|^{2+\delta}]}{\left(\sum_{v=-2L}^{2L} \sum_{u=0}^{2^\ell-1} (\alpha_v 2^{-\ell})^2 \mathbb{E}[|\langle A, \log M_{J,u+(k_0+v)2^\ell,n} - \mathbb{E}[\log M_{J,u+(k_0+v)2^\ell,n}]\rangle|^2]\right)^{(2+\delta)/2}}.$$

In order to give an upper bound on this term, we use the uniform bounded moments condition (3.1), the lower bound $\sum_v \alpha_v^2 \geq 1/(4L+1)$ as well as the upper bound

$$\limsup_{n\to\infty} \max_{-2L\leq v\leq 2L} |\alpha_v| = \limsup_{J_0\to\infty} \lim_{J\to\infty} \max_{-2L\leq v\leq 2L} |\alpha_v| \leq \sup_{x\in[0,1]} \max_{i,j} \Phi_L(x)_{i,j} < \infty,$$

where $\Phi_L(x)$ is the limit matrix in (2.15); notice that the entries of $\Phi_L(x)$ are uniformly bounded due to the Hölder continuity of the fundamental solution φ_L, cf. Donoho [1993].

Combining these estimates, the term in question is of order $2^{-\delta\ell/2}$. Hence, the Lyapunov condition is satisfied. □

Appendix C: Technical results on the wild bootstrap in Section 4

Proof of Theorem 4.3. The proof is split in three parts. In the first part we study the limiting behavior of the conditional variance of the bootstrap estimator. In the second part, we verify then the Lindeberg condition. In the third part, we show the amendment. To facilitate the notation, we use the following abbreviations. Set $\ell := J - J_0$. Let $A \in Sym(d)$ and define for an arbitrary but fixed index k on scale J_0 the random variables

$$\begin{aligned}
\widehat{Z}_{u,n} &= \langle A,\ \log \widehat{M}_{J,2^\ell k+u,n}(J_0^*)\rangle,\\
Z_{u,n} &= \langle A, \log M_{J,2^\ell k+u,n}\rangle,\\
Z_u &= \langle A, \log M_{J,2^\ell k+u}\rangle
\end{aligned}$$

for $u \in \{0,\dots,2^\ell-1\}$. We choose $\delta,\gamma>0$ such that $p=\gamma(2+\delta)$, $\gamma\geq 2$ and such that δ is sufficiently small; see below for the admissible choices.

Part 1. Recall that $\log M^*_{J_0,k,n}(J_0^*) = 2^{-\ell}\sum_{u=0}^{2^\ell-1} \log M^*_{J,u+k2^\ell,n}(J_0^*)$ by the midpoint pyramid algorithm. Hence the conditional variance of this element is

$$\begin{aligned}
\mathrm{Var}^*(2^{\ell/2}\langle A, \log M^*_{J_0,k,n}\rangle) &= 2^{-\ell}\sum_{u=0}^{2^\ell-1} \mathrm{Var}^*\langle A, \log \widehat{M}_{J,u+k2^\ell,n}(J_0^*) + \varepsilon^*_{J,u+k2^\ell,n}\rangle\\
&= 2^{-\ell}\sum_{u=0}^{2^\ell-1} \mathbb{E}^*\langle A, \widehat{\varepsilon}_{J,u+k2^\ell,n} V_{J,u+k2^\ell,n}\rangle^2\\
&= 2^{-\ell}\sum_{u=0}^{2^\ell-1} (\widehat{Z}_{u,n} - Z_{u,n})^2.
\end{aligned}$$

We split each summand in a main term and remainder terms as follows

$$\widehat{Z}_{u,n} - Z_{u,n} = (\widehat{Z}_{u,n} - \mathbb{E}[\widehat{Z}_{u,n}]) + (\mathbb{E}[\widehat{Z}_{u,n}] - Z_u) + (Z_u - Z_{u,n}). \tag{C.1}$$

We begin with the term $2^{-\ell}\sum_{u=0}^{2^\ell-1}(\widehat{Z}_{u,n} - \mathbb{E}[\widehat{Z}_{u,n}])^2$ on which we apply the following simple consequence of the Burkholder inequality; we state this consequence as a lemma:

Lemma C.1. *Let $W_1, \dots, W_n$ be independent, real-valued, centered random variables such that $\max_i \mathbb{E}[|W_i|^q]^{1/q} \le m$, for some $q \ge 2$ and an $m \in \mathbb{R}_+$. Then $\mathbb{E}[|\sum_i W_i|^q] \le C_q^q\, m^q\, n^{q/2}$ for a certain $C_q \in \mathbb{R}_+$.*

Proof of Lemma C.1. Using the Burkholder inequality, $\mathbb{E}[|W_i|^q] \le C_q^q \mathbb{E}[(\sum_i W_i^2)^{q/2}]$ for some $C_q \in \mathbb{R}_+$. As $q \ge 2$, we apply the Minkowski inequality to obtain the result. □

Utilizing (2.15), we find with Lemma C.1 and the moment condition on $Z_{u,n}$ from (3.1) as well as the representation of the wavelet estimator from (B.5) that for each $\gamma, \varepsilon > 0$

$$\mathbb{P}\Big(2^{-\ell} \sum_{u=0}^{2^\ell-1} (\widehat{Z}_{u,n} - \mathbb{E}[\widehat{Z}_{u,n}])^2 \ge \varepsilon\Big) \le \sum_{u=0}^{2^\ell-1} \mathbb{P}((\widehat{Z}_{u,n} - \mathbb{E}[\widehat{Z}_{u,n}])^2 \ge \varepsilon)$$
$$\le \varepsilon^{-\gamma} \sum_{u=0}^{2^\ell-1} \mathbb{E}[|\widehat{Z}_{u,n} - \mathbb{E}[\widehat{Z}_{u,n}]|^{2\gamma}] \le C_1 \varepsilon^{-\gamma}\, 2^{(J-J_0)-(J-J_0^*)\gamma},$$

for a constant C_1, which does not depend on n. (Note that in the first inequality, we need to use this rough estimate because the $(\widehat{Z}_{u,n})_u$ are dependent.) Recall that J, J_0 and J_0^* are increasing with n such that $J - J_0 \to \infty$ respectively $J - J_0^* \to \infty$ and hence, provided $\delta > 0$ is sufficiently small, $\sum_n 2^{(J-J_0)-(J-J_0^*)p/(2+\delta)} < \infty$ and we can apply the Borel-Cantelli Lemma to conclude that the partial sum

$$2^{-\ell} \sum_{u=0}^{2^\ell-1} (\widehat{Z}_{u,n} - \mathbb{E}[\widehat{Z}_{u,n}])^2 \to 0 \quad a.s., \quad n \to \infty.$$

Moreover, there is a constant $C_2 \in \mathbb{R}_+$ (independent of n) such that

$$(\mathbb{E}[\widehat{Z}_{u,n}] - Z_u)^2 \le C_2 2^{-2NJ_0^*}$$

uniformly in $u \in \{0, \dots, 2^\ell - 1\}$ by the approximation result from [Chau, 2018, Appendix A]. Thus, the deterministic sum $2^{-\ell} \sum_{u=0}^{2^\ell-1} (\mathbb{E}[\widehat{Z}_{u,n}] - Z_u)^2$ converges to 0 as $n \to \infty$.

Finally, observe that $2^{-\ell} \sum_{u=0}^{2^\ell-1} (Z_{u,n} - Z_u)^2$ has expectation

$$2^{-\ell} \sum_{u=0}^{2^\ell-1} \langle A, \mathcal{C}(2^{-J_0}k + 2^{-J}u)A\rangle.$$

Hence, using Lemma C.1 a second time, for each $\gamma > 0$ there is a $C_3 \in \mathbb{R}_+$ (independent of n) such that

$$\mathbb{P}\Big(2^{-\ell} \Big| \sum_{u=0}^{2^\ell-1} (Z_{u,n} - Z_u)^2 - \mathbb{E}[(Z_{u,n} - Z_u)^2]\Big| \ge \varepsilon\Big)$$

$$\leq 2^{-\ell\gamma}\varepsilon^{-\gamma}\ \mathbb{E}\Big[\Big|\sum_{u=0}^{2^\ell-1}(Z_{u,n}-Z_u)^2-\mathbb{E}[(Z_{u,n}-Z_u)^2]\Big|^\gamma\Big]\leq C_3\varepsilon^{-\gamma}2^{-(J-J_0)\gamma/2}.$$

Consequently, provided $\sum_n 2^{-(J-J_0)p/(2(2+\delta))}<\infty$, which is satisfied if we choose δ sufficiently small, we have

$$2^{-\ell}\sum_{u=0}^{2^\ell-1}(Z_{u,n}-Z_u)^2-\mathbb{E}[(Z_{u,n}-Z_u)^2]\to 0\quad a.s.\quad (n\to\infty).$$

Thus, using the convergence result of the covariance operator from (3.13)

$$2^{-\ell}\sum_{u=0}^{2^\ell-1}(Z_{u,n}-Z_u)^2\to 2^{J_0}\int_{k2^{-J_0}}^{(k+1)2^{-J_0}}\langle A,\mathcal{C}(u)A\rangle\,\mathrm{d}u>0\quad a.s.$$

In particular, $\mathrm{Var}^*(2^{\ell/2}\langle A,\log M^*_{J_0,k,n}(J_0^*)\rangle)/\mathrm{Var}(2^{\ell/2}\langle A,\log M_{J_0,k,n}\rangle)\to 1$ with probability 1 as $n\to\infty$. Consequently, we have with probability 1 as $n\to\infty$ that

$$\mathrm{Var}^*(2^{\ell/2}\langle A,\log \widehat{M}^*_{J,k,n}(J_0)\rangle)/\mathrm{Var}(2^{\ell/2}\langle A,\log \widehat{M}_{J,k,n}\rangle)\to 1.$$

Part 2. For simplicity, we verify the Lindeberg condition only for $2^{\ell/2}\log M^*_{J_0,k,n}(J_0^*)=2^{-\ell/2}\sum_{u=0}^{2^\ell-1}\log M^*_{J,u+2^\ell k,n}(J_0^*)$; the actual verification for the wavelet estimator works in the same fashion but involves a little more notation.

Note that $\mathbb{E}^*[\log M^*_{J,k,n}(J_0^*)]=\log\widehat{M}_{J,k,n}(J_0^*)$ and that $\log M^*_{J,k,n}(J_0^*)-\log\widehat{M}_{J,k,n}(J_0^*)=\varepsilon^*_{J,k,n}(J_0^*)$. Moreover, as the conditional variance of this random variable converges $a.s.$ by the previous arguments, it is sufficient to prove the following Lindeberg condition in order to verify the conditional CLT

$$L_n^*(\mu)=\sum_{u=0}^{2^\ell-1}\mathbb{E}^*\Big[|2^{-\ell/2}\langle A,\varepsilon^*_{J,u+2^\ell k,n}(J_0^*)\rangle|^2\mathbb{1}\{|2^{-\ell/2}\langle A,\varepsilon^*_{J,u+2^\ell k,n}(J_0^*)\rangle|>\mu\}\Big]\to 0$$

for each $\mu>0$ with probability 1. Let $\delta>0$, then we have

$$L_n^*(\mu)\leq\mu^{-\delta}\ \mathbb{E}^*[|V_{0,0,n}|^{2+\delta}]\ 2^{-\ell(1+\delta/2)}\sum_{u=0}^{2^\ell-1}|\langle A,\widehat{\varepsilon}_{J,u+2^\ell k,n}(J_0^*)\rangle|^{2+\delta}.$$

Since $\langle A,\widehat{\varepsilon}_{J,u+2^\ell k,n}(J_0^*)\rangle=\widehat{Z}_{u,n}-Z_{u,n}$, we rely on the same decomposition as in (C.1). Similarly as in the first part of the proof but this time with the exponent $2+\delta$ instead of 2, we can show that

$$2^{-\ell}\sum_{u=0}^{2^\ell-1}|\widehat{Z}_{u,n}-\mathbb{E}[\widehat{Z}_{u,n}]|^{2+\delta}=o_{a.s.}(1)$$

for the choices of δ and γ. Furthermore, for the choices of δ and γ,

$$2^{-\ell}\sum_{u=0}^{2^\ell-1}|\mathbb{E}[\widehat{Z}_{u,n}]-Z_u|^{2+\delta}=o(1)\text{ and }2^{-\ell}\sum_{u=0}^{2^\ell-1}|Z_{u,n}-Z_u|^{2+\delta}=O_{a.s.}(1).$$

Using the monotonicity of $L_n^{boot}(\mu)$ in μ, this shows then that $L_n^{boot}(\mu) \to 0$ for $n \to \infty$ for all $\mu > 0$ with probability 1.

Part 3. The convergence in the Kolmogorov distance follows from a standard argument as the limiting Gaussian distribution of $\mathfrak{M}_{J,k,n}(J_0)$ and $\mathfrak{M}^*_{J,k,n}(J_0)$ has a continuous density on $Sym(d) \cong \mathbb{R}^{(d+1)d/2}$. Let $\mathcal{N}$ be a random variable having this Gaussian distribution and let $\mu > 0$. Let Γ be a finite grid on $Sym(d)$ with minimal element $\underline{x}$ and maximal element $\overline{x}$, i.e., $\underline{x} \le x \le \overline{x}$ for all $x \in \Gamma$. (Here $x \le y$ for $x, y \in Sym(d)$ if for each position (i,j), $x_{i,j} \le y_{i,j}$.) For $x \in Sym(d)$ with $\underline{x} \le x \le \overline{x}$, denote $\lfloor x \rfloor$ (resp. $\lceil x \rceil$) the elements in Γ which are closest to x in the maximum-norm and satisfy $\lfloor x \rfloor \le x$ (resp. $x \le \lceil x \rceil$). We choose Γ sufficiently dense in the sense that $\mathbb{P}(\mathcal{N} \le \underline{x}) \le \mu$, $1 - \mathbb{P}(\mathcal{N} \le \overline{x}) \le \mu$ and $\mathbb{P}(\mathcal{N} \le \lceil x \rceil) - \mathbb{P}(\mathcal{N} \le \lfloor x \rfloor) \le \mu$ for $\underline{x} \le x \le \overline{x}$.

Next, choose $N \in \mathbb{N}$ large enough such that $|\mathbb{P}(\mathfrak{M}_{J,k,n}(J_0) \le x) - \mathbb{P}(\mathcal{N} \le x)| \le \mu$ and $|\mathbb{P}(\mathfrak{M}^*_{J,k,n}(J_0) \le x) - \mathbb{P}(\mathcal{N} \le x)| \le \mu$ for all $x \in \Gamma$ for all $n \ge N$.

Then $|\mathbb{P}(\mathfrak{M}_{J,k,n}(J_0) \le x) - \mathbb{P}(\mathfrak{M}^*_{J,k,n}(J_0) \le x)| \le 4\mu$ for all $x \in Sym(d)$ for all $n \ge N$. □

Appendix D: Further numerical simulation results

Here we now complete our numerical simulations of Section 5 by depicting in Figure 7 first the wavelet estimator of curve $c_2(t)$ along with asymptotic and bootstrap confidence sets, based on $B = 100$ replicates and, for $J = 12$, with estimation parameters $J_0 = 7$ for the smoothing scale and $N = 5$. Note that J_0 has been chosen in order to undersmooth (as required for obtaining optimal confidence regions), which explains the somewhat wiggly behaviour of the curve estimator itself.

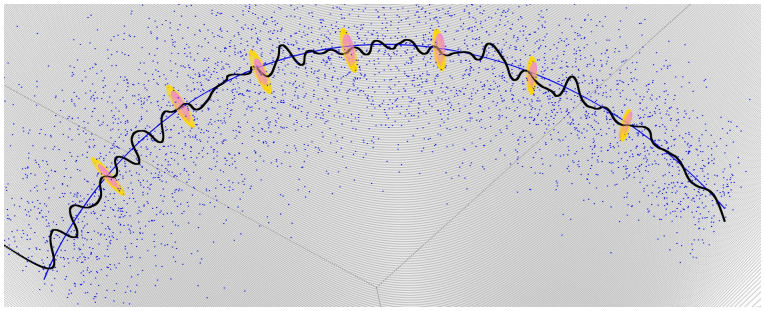


FIG 7. *AI-wavelet estimator (black), asymptotic (gold) and bootstrap (pink) 0.975-confidence sets for curve c_2 and $J = 12$, $J_0 = 7$.*

Then we proceed by showing in Figure 8 the same result for our curve $c_3(t)$.

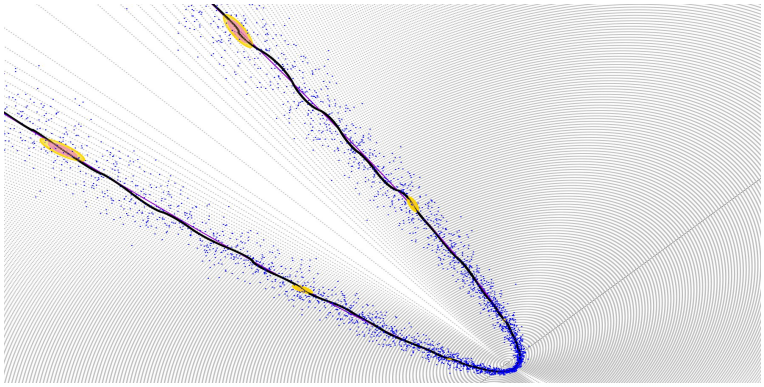


FIG 8. *AI-wavelet estimator (black), asymptotic (gold) and bootstrap (pink)* 0.975*-confidence sets for curve* c_3 *and* $J = 12$, $J_0 = 7$.

For completeness we show here below the same pointwise empirical coverage plots and histograms for curves $c_2(t)$ and $c_3(t)$ as they have been shown in Section 5 for curve $c_1(t)$.

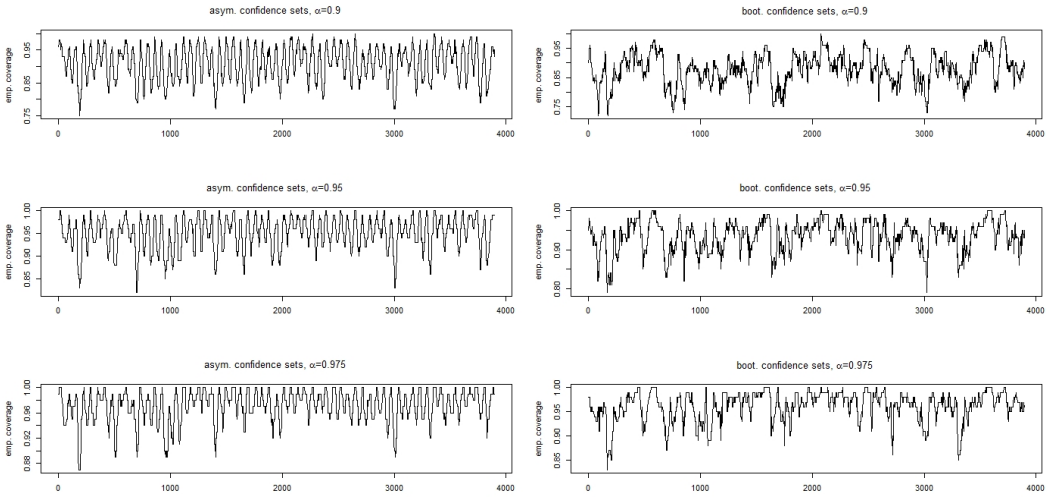


FIG 9. *Pointwise empirical coverage for curve* c_2 *and* $J = 12$, $J_0 = 7$, $N = 5$

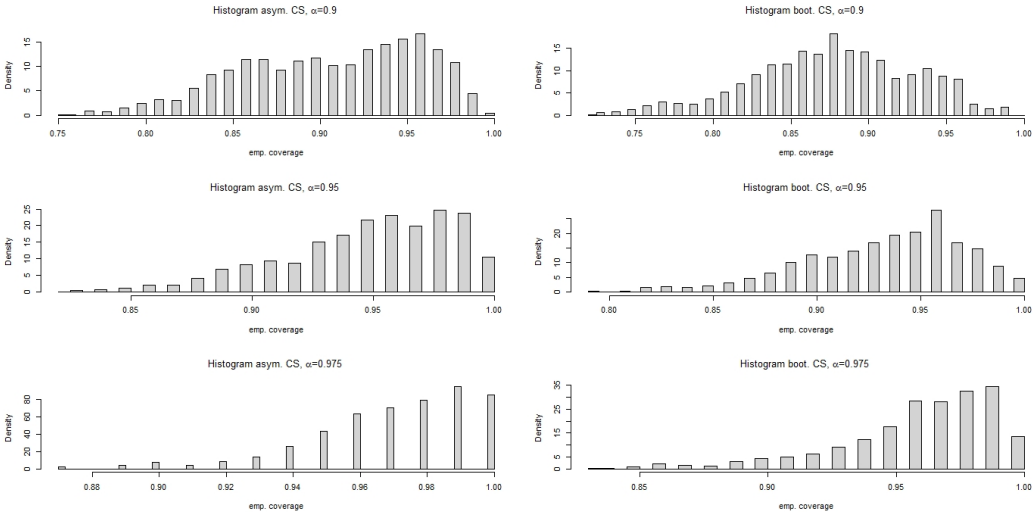


FIG 10. *Histogram of empirical coverage for curve c_2 and $J = 12$, $J_0 = 7$, $N = 5$*

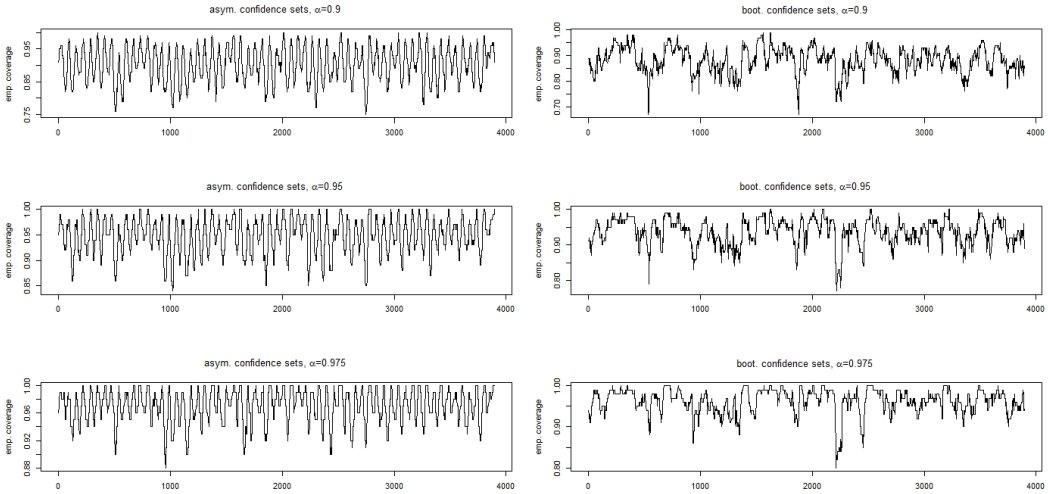


FIG 11. *Pointwise empirical coverage for curve c_3 and $J = 12$, $J_0 = 7$, $N = 5$*

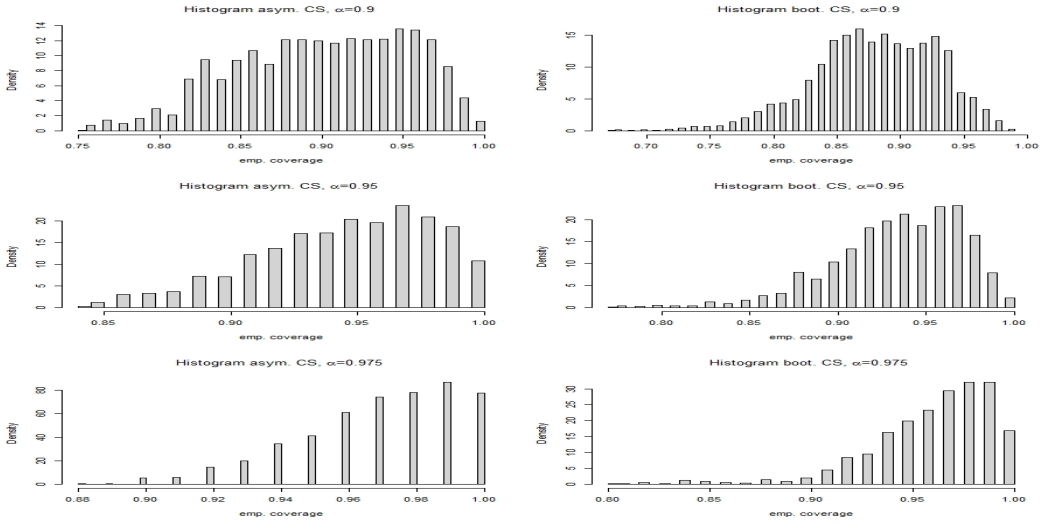


FIG 12. *Histogram of empirical coverage for curve* c_3 *and* $J = 12$, $J_0 = 7$, $N = 5$

To illustrate the typical bias effect of confidence regions based on nonparametric curve estimators for a sample size of order 10^3 (i.e. for the choice of $J = 10$) we finally show by Tables 5 and 6, for curve $c_1(t)$, a comparison of the average empirical coverages for regions centred around the target curve with those centred around the mean of the wavelet estimator which is approximated - at each considered point in time - by the sample mean over 50 repetitions.

TABLE 5
Simulation results for curve c_1 *and* $J = 10$, $J_0 = 6$, $N = 5$. *Scaled averaged volumes are given in units of* 10^{-4}.

	alpha=0.9		alpha=0.95		alpha=0.975	
	Asym.	Boot.	Asym.	Boot.	Asym.	Boot.
Emp. Cov.	0.5204	0.8748	0.5936	0.9344	0.6528	0.9616
Scaled Ave. Vol.	0.055	0.240	0.076	0.384	0.100	0.550

TABLE 6
Simulation results for curve c_1 *and* $J = 10$, $J_0 = 6$, $N = 5$, *mean centred. Scaled averaged volumes are given in units of* 10^{-4}.

	alpha=0.9		alpha=0.95		alpha=0.975	
	Asym.	Boot.	Asym.	Boot.	Asym.	Boot.
Emp. Cov.	0.9041	0.8673	0.9480	0.9261	0.9698	0.9551
Scaled Ave. Vol.	0.055	0.194	0.076	0.318	0.100	0.462

Appendix E: The log-Euclidean Metric

For the readers ease and clarity in notation, we summarize the results of Arsigny et al. [2007] necessary for this work. We also add the formula for parallel transport under the log-Euclidean metric. For general treatments of modern differential and Riemannian geometry we refer to Lee [1997] and Tu [2017].

E.1. Preliminaries

Theorem E.1 (cf. Arsigny et al. [2007], Theorem 2.2)**.** *The matrix exponential* $\exp\colon M(d) \to GL(d)$ *is a* C^∞*-mapping and its differential* $d\exp\colon TM(d) \cong M(d) \to M(d) \cong TGL(d)$ *is given by*

$$d_A \exp\colon T_A M(d) \cong M(d) \to M(d) \cong T_{\exp(A)} GL(d),$$
$$B \mapsto \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{l=0}^{k-1} A^l B A^{k-l-1}.$$

Corollary E.2 (cf. Arsigny et al. [2007], Cor. 2.3)**.** *In particular* $d_0 \exp = \mathrm{Id}\colon M(d) \to M(d)$ *and*

$$\operatorname{trace}(d_A \exp(B)) = \operatorname{trace}(\exp(A)B), \quad A, B \in M(d).$$

Definition E.3 (matrix logarithm). A matrix $B \in M(d)$ is called a logarithm of a matrix $A \in GL(d)$ if $\exp(B) = A$.

Remark E.4 (cf. Arsigny et al. [2007]). Since the mapping $\exp\colon M(d) \to GL(d)$ is *not surjective* the logarithm of a matrix may not exist in general. However, if $A \in GL(d)$ has no (complex) eigenvalues on the (closed) negative real line, then A has a *unique* real logarithm whose (complex) eigenvalues have an imaginary part in $(-\pi, \pi)$. This particular logarithm is called *principal logarithm* and will be denoted $\log(A)$ whenever it is defined. Especially $\log(A)$ is defined for any $A \in Sym^+(d)$ and is symmetric.

E.2. Log-Euclidean Metric(s) on $Sym^+(d)$

The following theorem summarizes Arsigny et al. [2007] Theorem 2.6, Proposition 2.7 and Theorem 2.8.

Theorem E.5. *The mappings* $\exp\colon Sym(d) \to Sym^+(d)$ *and* $\log\colon Sym^+(d) \to Sym(d)$ *are* C^∞ *and one-to-one, i.e., the spaces are diffeomorphic and* $\exp^{-1} = \log$*. Also*

$$d_A \exp\colon T_A Sym(d) \cong Sym(d) \to Sym(d) \cong T_{exp(A)} Sym^+(d)$$

is invertible for all $A \in Sym(d)$*. Topologically* $Sym^+(d)$ *is an open convex half-cone of* $Sym(d)$ *and therefore a submanifold of* $Sym(d)$*.*

The idea for the construction of log-Euclidean metrics is to use the matrix exponential to transport the additive group structure of $Sym(d)$ to $Sym^+(d)$. To this end define the *logarithmic product* $\odot\colon Sym^+(d)\times Sym^+(d)\to Sym^+(d)$ by

$$S_1\odot S_2 := \exp(\log(S_1)+\log(S_2)).$$

Theorem E.6. *$(Sym^+(d),\odot)$ is a commutative Lie group. The neutral element is the identity matrix and the inverse element is simply the matrix inverse. Furthermore the matrix exponential*

$$\exp\colon (Sym(d),+)\to(Sym^+(d),\odot)$$

is a Lie group isomorphism.

In particular one-parameter subgroups of $(Sym^+(d),\odot)$ are of the form $\exp(tV)_t$, where $(tV)_t$, $V\in Sym(d)$ are simply the one-parameter subgroups of $(Sym(d),+)$. Also the Lie group exponential

$$\exp\colon T_{\mathrm{Id}}Sym^+(d)=Sym(d)\to Sym^+(d)$$

is just the ordinary matrix exponential.

Proof. Arsigny et al. [2007] Proposition 3.2, Theorem 3.3 and Proposition 3.4 □

Now any inner product $\langle\cdot,\cdot\rangle_{\mathrm{Id}}$ on the vector space $Sym(d)\cong T_{\mathrm{Id}}Sym^+(d)$ can be extended to a *Riemannian metric* g on the whole manifold $(Sym^+(d),\odot)$ by

$$\begin{aligned} g_S(U,V)=\langle U,V\rangle_S :=& \langle (d_{\mathrm{Id}}l_S)^{-1}U,(d_{\mathrm{Id}}l_S)^{-1}V\rangle_{\mathrm{Id}}, \\ =& \langle (d_{\mathrm{Id}}r_S)^{-1}U,(d_{\mathrm{Id}}r_S)^{-1}V\rangle_{\mathrm{Id}}, \end{aligned} \tag{E.1}$$

because $U,V\in T_SSym^+(d)\cong Sym(d)$, $S\in Sym^+(d)$ and where

$$\begin{aligned} l_S\colon Sym^+(d)\to Sym^+(d),\quad W\mapsto S\odot W \quad (\textit{left translation})\\ d_{\mathrm{Id}}l_S\colon T_{\mathrm{Id}}Sym^+(d)\to T_SSym^+(d). \end{aligned}$$

Since the logarithmic product $\odot$ is commutative, we have $l_S=r_S$, where

$$r_S\colon Sym^+(d)\to Sym^+(d),\quad W\mapsto W\odot S$$

denotes the *right translation*. Therefore we have

Corollary E.7 (cf. Arsigny et al. [2007], Cor. 3.7 and Cor. 3.10)**.** *Any Riemannian metric obtained by (E.1) is a bi-invariant metric on $(Sym^+(d),\odot)$, i.e.*

$$\begin{aligned} \langle U,V\rangle_{\mathrm{Id}} &= \langle d_{\mathrm{Id}}l_S(U),d_{\mathrm{Id}}l_S(V)\rangle_S && \textit{(left-invariant)}\\ &\big(=\langle d_{\mathrm{Id}}r_S(U),d_{\mathrm{Id}}r_S(V)\rangle_S\big). && \textit{(right-invariant)} \end{aligned}$$

Also endowed with such a metric $Sym^+(d)$ becomes a flat Riemannian manifold, i.e., the curvature tensor (resp. the sectional curvature) is null.

Definition E.8 (log-Euclidean metric, cf. Arsigny et al. [2007], Def. 3.8). Any bi-invariant metric on $(Sym^+(d),\odot)$ is called log-Euclidean metric.

E.3. Geometry of $Sym^+(d)$ under log-Euclidean Metric(es)

Denote by g a bi-invariant metric on $(Sym^+(d), \odot)$ obtained by (E.1). Recall that

$$\frac{d}{dt}\exp(S+tV)|_{t=0} = d_S\exp(V) \in T_{\exp(S)}Sym^+(d), \quad V \in T_S Sym(d)$$

Define $\gamma_S(\cdot, L)\colon [0,1] \to Sym^+(d)$ for $S \in Sym^+(d)$ und $L \in T_S Sym^+(d)$ by

$$\gamma_S(t; L) = \exp\big(\log(S) + t(d_{\log(S)}\exp)^{-1}L\big),$$

then $\gamma_S(0; L) = S$ and $\dot\gamma_S(0; L) = L$. Since $\gamma_S(\cdot; L)$ is a one-parameter subgroup of $(Sym^+(d), \odot)$ (cf. Theorem E.6) by Theorem 3.6 in Arsigny et al. [2007] $\gamma_S(\cdot; L)$ is a geodesic (with resp. to g) starting in S in direction L. Therefore the *exponential map* $\mathrm{Exp}_S\colon T_S Sym^+(d) \to Sym^+(d)$ is given by

$$\mathrm{Exp}_S(L) := \gamma_S(1; L) = \exp\big(\log(S) + (d_{\log(S)}\exp)^{-1}L\big).$$

Now consider $\log\circ\exp = \mathrm{Id}\colon Sym^+(d) \to Sym^+(d)$, then

$$\begin{aligned} d_S(\log\circ\exp)(U) &= d_{\exp(S)}\log\circ\, d_S\exp(U) \\ &= d_S\mathrm{Id}(U) = \frac{d}{dt}(S+tU)|_{t=0} = U \quad \forall U \in Sym^+(d) \\ \Leftrightarrow\quad & d_{\exp(S)}\log\circ\, d_S\exp = \mathrm{Id} \\ \Leftrightarrow\quad & d_{\exp(S)}\log = (d_S\exp)^{-1} \\ \Leftrightarrow\quad & d_S\log = (d_{\log(S)}\exp)^{-1}, \end{aligned}$$

which yields

$$\mathrm{Exp}_S(L) = \exp\big(\log(S) + d_S\log(L)\big). \tag{E.2}$$

Next put $\mathrm{Exp}_{S_1}(L) = S_2$ for $S_1, S_2 \in Sym^+(d)$. Solving (E.2) for $L \in T_{S_1}Sym^+(d)$ gives

$$\begin{aligned} \log(S_1) + d_{S_1}\log(L) = \log(S_2) \quad &\Leftrightarrow \quad d_{S_1}\log(L) = \log(S_2) - \log(S_1) \\ &\Leftrightarrow \quad L = d_{\log(S_1)}\exp\big(\log(S_2) - \log(S_1)\big). \end{aligned}$$

Therefore the *logarithmic map* $\mathrm{Log}_{S_1}\colon Sym^+(d) \to T_{S_1}Sym^+(d)$ is given by

$$\mathrm{Log}_{S_1}(S_2) = d_{\log(S_1)}\exp\big(\log(S_2) - \log(S_1)\big).$$

Furthermore for $\gamma(t) = \exp(tU)$, $U \in Sym(d)$ we have $\gamma(0) = \mathrm{Id}$ and $\dot\gamma(0) = U$, thus

$$\begin{aligned} d_{\mathrm{Id}}l_S(U) &= \frac{d}{dt}\, l_S\circ\gamma(t)|_{t=0} \\ &= \frac{d}{dt}\, S\odot\exp(tU)|_{t=0} = \frac{d}{dt}\exp(\log(S)+tU)|_{t=0} = d_{\log(S)}\exp(U) \end{aligned}$$

$$\Leftrightarrow \quad (d_{\mathrm{Id}} l_S)^{-1} = (d_{\log(S)} \exp)^{-1} = d_S \log .$$

The *log-Euclidean metric* g can consequently be explicitly written as

$$g_S(U, V) = \langle d_S \log(U), d_S \log(V) \rangle_{\mathrm{Id}}. \tag{E.3}$$

Accordingly for the *distance of two points* $S_1, S_2 \in Sym^+(d)$ we obtain

$$\begin{aligned} d^2(S_1, S_2) &= g^2_{S_1}(\mathrm{Log}_{S_1}(S_2), \mathrm{Log}_{S_1}(S_2)) \; (= \|\mathrm{Log}_{S_1}(S_2)\|^2_{S_1}) \\ &= \langle d_{S_1} \log(\mathrm{Log}_{S_1}(S_2)), d_{S_1} \log(\mathrm{Log}_{S_1}(S_2)) \rangle_{\mathrm{Id}} \\ &= \| \log(S_2) - \log(S_1) \|^2_{\mathrm{Id}}. \end{aligned}$$

To conclude the discussion we note that Theorem 3.6 in Arsigny et al. [2007] together with Theorem E.6 immediately imply that the unique geodesic from S_1 to S_2 in $(Sym^+(d), \odot)$ is given by

$$\gamma(t; S_1, S_2) = \exp((1-t)\log(S_1) + t \log(S_2)), \quad t \in [0, 1]. \tag{E.4}$$

Also (E.3) shows that $\log : (Sym^+(d), g) \to (Sym(d), \langle \cdot, \cdot \rangle_{\mathrm{Id}})$ is an *isometry* (and likewise $\exp\colon (Sym(d), \langle \cdot, \cdot \rangle_{\mathrm{Id}}) \to (Sym^+(d), g)$), so the following diagram commutes

$$\begin{array}{ccc} T_{S_1} Sym^+(d) & \xrightarrow{\Gamma^{S_2}_{S_1}} & T_{S_2} Sym^+(d) \\ \Big\downarrow d_{S_1} \log & & \Big\downarrow d_{S_2} \log \\ T_{\log(S_1)} Sym(d) & \xrightarrow{\widetilde{\Gamma}^{\log(S_2)}_{\log(S_1)}} & T_{\log(S_2)} Sym(d) \end{array}$$

where $\Gamma^{S_2}_{S_1}$ denotes the *parallel transport* (in $(Sym^+(d), g)$) along the geodesic $\gamma(\cdot; S_1, S_2)$. The corresponding parallel transport $\widetilde{\Gamma}^{\log(S_2)}_{\log(S_1)}$ in $Sym(d)$ (along $\log \circ \gamma$) is just the identity map since $Sym(d)$ is a vector space. Therefore for $U \in T_{S_1} Sym^+(d)$ we have

$$\begin{aligned} \Gamma^{S_2}_{S_1}(U) &= (d_{S_2} \log)^{-1} \circ d_{S_1} \log(U) \\ &= d_{\log(S_2)} \exp \big(d_{S_1} \log(U) \big). \end{aligned}$$

In particular the parallel transport of $U \in T_S Sym^+(d)$ to the tangent space at the identity, $T_{\mathrm{Id}} Sym^+(d) = Sym(d)$, is

$$\Gamma^{\mathrm{Id}}_S(U) = d_S \log(U)$$

since $d_{\log(\mathrm{Id})} \exp = d_0 \exp = \mathrm{Id}$, cf. Yuan et al. [2012] Equation (14).

Finally, by choosing $\langle U, V \rangle_{\mathrm{Id}} = \langle U, F \rangle_F = \mathrm{trace}(UV)$ the corresponding log-Euclidean metric is invariant under orthogonal similarity transformations.

Lemma E.9. *For $O \in GL(d)$ orthogonal and $S_1, S_2 \in Sym^+(d)$ we have*

$$d(OS_1O^T, OS_2O^T) = d(S_1, S_2).$$

Proof. Simply observe that $\log(ASA^{-1}) = A\log(S)A^{-1}$ for any non-singular matrix $A \in M(d)$ and $S \in Sym^+(d)$. Hence

$$\begin{aligned} d(OS_1O^T, OS_2O^T) &= \|O(\log S_1 - \log S_2)O^T\|_F \\ &= \|\log S_1 - \log S_2\|_F = d(S_1, S_2). \end{aligned}$$

□

We close this section by summarizing the derived geometric tools of the log-Euclidean metric as follows (recall that $d_S \log$ denotes the differential of the matrix logarithm at a point $S \in Sym^+(d)$):

$Sym^+(d)$	manifold
$T_S Sym^+(d) = S \times Sym(d) \cong Sym(d)$	tangent space
$g_S(U,V) = \operatorname{trace}(d_S \log(U) d_S \log(V)), \quad U, V \in T_S Sym^+(d)$	log-Eucl. metric
$d(S_1, S_2) = \|\log(S_2) - \log(S_1)\|_F, \quad S_1, S_2 \in Sym^+(d)$	log-Eucl. distance
$\gamma(t; S_1, S_2) = \exp((1-t)\log(S_1) + t\log(S_2)), \quad S_1, S_2 \in Sym^+(d)$	geodesic
$\operatorname{Exp}_S(U) = \exp\big(\log(S) + d_S \log(U)\big), \quad S \in Sym^+(d),\ U \in T_S Sym^+(d)$	exponential map
$\operatorname{Log}_{S_1}(S_2) = d_{\log(S_1)} \exp\big(\log(S_2) - \log(S_1)\big), \quad S_1, S_2 \in Sym^+(d)$	logarithmic map
$\Gamma_{S_1}^{S_2}(U) = d_{\log(S_2)} \exp\big(d_{S_1} \log(U)\big), \quad S_1, S_2 \in Sym^+(d),\ U \in T_{S_1} Sym^+(d)$	parallel transport
$\Gamma_S^{\mathrm{Id}}(U) = d_S \log(U), \quad S \in Sym^+(d),\ U \in T_S Sym^+(d)$	

TABLE 7
Abbreviations for the relevant differential geometric tools