Low-rank multi-parametric covariance identification

Low-rank multi-parametric covariance identification

Antoni Musolas · **Estelle Massart** · **Julien M. Hendrickx** · **P.-A. Absil** · **Youssef Marzouk**

Abstract We propose a differential geometric approach for building families of low-rank covariance matrices, via interpolation on low-rank matrix manifolds. In contrast with standard parametric covariance classes, these families offer significant flexibility for problem-specific tailoring via the choice of "anchor" matrices for interpolation, for instance over a grid of relevant conditions describing the underlying stochastic process. The interpolation is computationally tractable in high dimensions, as it only involves manipulations of low-rank matrix factors. We also consider the problem of covariance identification, i.e., selecting the most representative member of the covariance family given a data set. In this setting, standard procedures such as maximum likelihood estimation are nontrivial because the covariance family is rank-deficient; we resolve this issue by casting the identification problem as distance minimization. We demonstrate the utility of these differential geometric families for interpolation and identification in a practical application: wind field covariance approximation for unmanned aerial vehicle navigation.

Keywords covariance approximation · interpolation ~~optimization~~ on manifolds · positive-semidefinite matrices · Riemannian metric · geodesic · low-rank covariance function · maximum likelihood

Mathematics Subject Classification (2000) 53C22 · 62J10

1 Introduction

One of the most fundamental problems in multivariate analysis and uncertainty quantification is the construction of covariance matrices. Covariance matrices are an essential tool in climatology, econometrics,

Antoni Musolas and Youssef Marzouk
Center for Computational Science & Engineering, Massachusetts Institute of Technology, Cambridge, MA USA
E-mail: musolas@mit.edu

Estelle Massart
Mathematical Institute, University of Oxford, OX2 6GG, UK and National Physical Laboratory, Hampton Road, Teddington, Middlesex, TW11 0LW, UK.

Julien M. Hendrickx and P.-A. Absil
ICTEAM Institute, UCLouvain, Avenue Georges Lemaître 4 bte L4.05.01, 1348 Louvain-la-Neuve, Belgium

model order reduction, biostatistics, signal processing, and geostatistics, among other applications. As a specific example (which we shall revisit in this paper), covariance matrices of wind velocity fields [35, 36,43,8] capture the relationships among wind velocity components at different points in space; modeling these relationships is essential to incorporating pointwise measurements of the velocity field into the overall wind model. ~~These relationships enable recursive approximation or updating of the wind field as new pointwise measurements become available.~~ Similarly, in oil reservoir modeling [49,50], covariance matrices allow information from borehole measurements to be propagated into more accurate global estimates of the permeability field. In telecommunications [42], covariance matrices and their eigenvectors are paramount for discerning between signal and noise.

Widely used methods in spatial statistics [14,54,62] include variogram estimation [13,11] (often a first step in kriging [61,17,25]) or tapering of sample covariance matrices [24,19]. Other regularized covariance estimation methodologies include sparse covariance and precision matrix estimation [6,18,7] and many varieties of shrinkage [37,38,59,39]. Many of these methods make rather generic assumptions on the covariance matrix (e.g., sparsity of the precision, or some structure in the spectrum); others (e.g., Matérn covariance kernels or particular variogram models) assume that the covariance is described within some parametric family. The latter can make the estimation problem quite tractable, even with relatively limited data, since the number of parameters ~~degrees of freedom~~ in the family may be small.

Unfortunately, standard parametric covariance families (e.g., Matérn) [12,53,56] can be insufficiently flexible for practical applications. For instance, in wind field modeling, these covariance families are not expressive enough to capture the non-stationary and multiscale features of the velocity or vorticity [32, 33,34]. Nonparametric approaches such as sparse precision estimation may be less restrictive, but neither approach easily allows prior knowledge—such as known wind covariances at other conditions—to be incorporated. Moreover, most of these methods yield full-rank covariance matrices, which are impractical for high-dimensional problems. For example, direct manipulation of full-rank covariances in high-dimensional settings might preclude recursive estimation (i.e., conditioning) from being performed online. Also, many applications naturally produce covariances that are well approximated by low-rank matrices; as explained in Section 5, this is the case for the above-mentioned wind field modeling problem.

In this paper, we build low-rank covariance families using interpolation techniques on matrix manifolds, whose development has been an active research topic in the last few years; see, e.g., [5,57,26,31, 9]. We propose to rely on piecewise-Bézier curves and surfaces on manifolds that have been investigated, e.g., in [2,3,21,47]. As for the space on which to interpolate, an immediate choice would have been the set of all $n \times n$ covariance matrices, namely the set of $n \times n$ symmetric positive-semidefinite (PSD) matrices; however this set is not a Riemannian manifold. (Specifically, it is not a submanifold of $\mathbb{R}^{n \times n}$.) The above-mentioned techniques, which in general rely on successive evaluations of the Riemannian exponential and logarithm, are therefore not applicable. Instead, we will consider the set of $n \times n$ covariance matrices of fixed rank r, which is known to be a manifold; see, e.g., [64]. In the wind field application and others, truncating the rank of the covariance matrices to some common value, before interpolation, has no significant impact on the accuracy of the resulting approximations. Hence, throughout this work, we assume the matrices to belong to the manifold of fixed-rank positive-semidefinite matrices.

Interpolation on positive-(semi)definite matrix manifolds has been pioneered and popularized by, among others, [52] and [46], who both investigated the full-rank case ($r = n$). Recently, [48] proposed to use interpolation on the manifold of positive-definite matrices to generate families of covariance matrices, in the same spirit as this work. Extension to the low-rank case is challenging, because of the lack of a preferred metric on the manifold of fixed-rank positive-semidefinite matrices: unlike the full-rank case, to our knowledge, no metric achieves the desirable property of turning the manifold into a complete metric space with practical closed-form expressions for the Riemannian exponential and logarithm. Among the proposed geometries (see [4,64,65]), we resort to the geometry proposed in [28] and further developed in [44], as it has the crucial advantage of providing a closed-form expression for the endpoint geodesic

problem (i.e., the Riemannian logarithm), which appears as a pervasive subproblem in Bézier-type interpolation techniques on manifolds. We point out that univariate interpolation on the manifold of fixed-rank PSD matrices has already been applied, e.g., to the wind field approximation problem in [22], to protein conformation in [41], to computer vision and facial recognition in [29, 63], and to parametric model order reduction in [45]. The present work is, to our knowledge, a first step towards multivariate interpolation of fixed-rank positive-semidefinite matrices. Although we are restricting the scope to Bézier surfaces on manifolds, we emphasize that other multivariate interpolation strategies exist, e.g., Gaussian radial basis function (RBF)-based positive-definite kernels on manifolds in [27], quasi-interpolation operators in a Riemannian manifold in [23], and geodesic finite elements to obtain spaces of high order approximation in [58]. Application of these methods to the manifold of fixed-rank positive-semidefinite matrices could be the focus of future research.

There are three original contributions in this paper. First, we devise new *low-rank* parameterized covariance families, by interpolating a set of problem-specific anchor covariance matrices. The rank of the anchor matrices is assumed (or truncated) to be some value r, usually much smaller than the size n of the matrices. The resulting covariance families are shown to contain only matrices of rank less than or equal to r. In high-dimensional applications, this reduces the computational cost of interpolation, since only low-rank factors of the covariance matrices need to be manipulated. Second, to identify the most representative member of a covariance family given some data, we minimize an appropriate loss function, as an alternative to maximum likelihood estimation. Indeed, when the covariance family is low-rank, maximum likelihood estimation is not trivial, as the probability density of the data (assumed Gaussian) is degenerate. Observe that both the definition of the covariance family and the identification problem do not require an appeal to asymptotic statistical properties, which are not investigated here. Third, we demonstrate the accuracy and efficacy of these covariance families on an application: wind field velocity characterization for UAV navigation.

The rest of this paper is organized as follows. We summarize the tools needed to work on the manifold of fixed-rank PSD matrices in Section 2. We introduce our new covariance functions in Section 3, for both the one- and the multi-parameter cases. In Section 4, we present methods to solve the covariance identification problem via distance minimization. In Section 5, we illustrate the behaviors of the different covariance functions on a case study: wind field approximation.

2 The geometry of the set of positive-semidefinite matrices

In this section, we define useful tools to work on the manifold $\mathscr{S}_+(r,n)$ of positive-semidefinite (PSD) matrices, with rank r and size $n \times n$, with $r < n$. The interpolation algorithms used in this paper mostly rely on expressions for the geodesics, and more precisely, the Riemannian exponential (geodesic with given initial conditions) and the Riemannian logarithm (endpoint geodesic problem). Before discussing those, we briefly describe the geometry and metric considered.

Several metrics have been proposed for this manifold but, to our knowledge, none of them manages to turn it into a complete metric space with a closed-form expression for endpoint geodesics. We use the metric naturally inherited from the quotient geometry $\mathscr{S}_+(r,n) \simeq \mathbb{R}_*^{n \times r}/\mathscr{O}(r)$ proposed in [28] and further developed in [44], with $\mathbb{R}_*^{n \times r}$ endowed with the Euclidean metric. This geometry relies on the fact that a matrix $A \in \mathscr{S}_+(r,n)$ can be factorized as $A = Y_A Y_A^\top$, where the factor $Y_A \in \mathbb{R}_*^{n \times r}$ has full column rank. The decomposition is not unique, as each factor $\tilde{Y}_A := Y_A Q$, with $Q \in \mathscr{O}(r)$ an orthogonal matrix, leads to the same product. As a consequence, any PSD matrix A is represented by an equivalence class:

$$[Y_A] = \{Y_A Q | Q \in \mathscr{O}(r)\},$$

where $Y_A \in \mathbb{R}_*^{n \times r}$ is an arbitrary matrix satisfying $Y_A Y_A^T = A$. In our computations, we work with representatives of the equivalence classes. For example, the geodesic between two PSD matrices A and B will

be computed based on two arbitrary representatives Y_A, Y_B, of the corresponding equivalence classes. The geodesic will of course be invariant to the choice of the representatives. Moreover, this approach saves computational cost as the representatives are of size $n \times r$, instead of $n \times n$.

The recent paper [44] provides, among others, expressions for minimizing geodesics on the quotient manifold $\mathbb{R}_*^{n \times r}/\mathscr{O}_r$. Since $\mathbb{R}_*^{n \times r}/\mathscr{O}_r$ is not complete, we cannot define geodesics between any arbitrarily chosen pair of points. We therefore make the following assumption throughout the paper:

Hypothesis 1 *The data matrices are such that all the geodesic segments to which we refer in the sequel are well-defined. In the specific case when we refer to a geodesic between two points $A, B \in \mathscr{S}_+(r, n)$ with corresponding equivalence class representatives $Y_A, Y_B \in \mathbb{R}_*^{n \times r}$ (i.e., an endpoint geodesic), this holds if and only if the matrix product $Y_A^\top Y_B$ is nonsingular.*

Hypothesis 1 is typically satisfied when the data matrices are sufficiently close to each other; see [44] for more information, and is less restrictive for some of the interpolation schemes discussed in this paper than for others. In particular, for the approaches relying on a section of the manifold (see below), one only requires the endpoint geodesics joining the data matrices to the reference point of the section to be well-defined. Hypothesis 1 is obviously more restrictive for the interpolation schemes involving combinations of geodesic segments (see Section 3.1.2 and Section 3.3.2), though we show numerically in Section 5 that these interpolation schemes work well in applications when no guarantee is known on the validity of Hypothesis 1, if we allow the geodesic segments to leave temporally the manifold. We finally mention that [2, §3.2] provides sufficient conditions relating the distance between the data matrices, the curvature, and the injectivity radius of the manifold, for Hypothesis 1 to hold in the case of the patchwise Bézier interpolation method described in Section 3.3.2, though these conditions are stricter than required for many applications.

The minimizing geodesic $\varphi_{A_1 \to A_2}$ between two PSD matrices A_1, A_2 satisfying Hypothesis 1, with representatives Y_{A_1} and Y_{A_2}, is given by:

$$\varphi_{A_1 \to A_2}(t) = Y_{\varphi_{A_1 \to A_2}(t)} Y_{\varphi_{A_1 \to A_2}(t)}^\top \quad \text{with} \quad Y_{\varphi_{A_1 \to A_2}(t)} = Y_{A_1} + t\dot{Y}_{A_1 \to A_2}. \tag{2.1}$$

In this expression, the vector $\dot{Y}_{A_1 \to A_2}$ is defined as $\dot{Y}_{A_1 \to A_2} = Y_{A_2} Q^\top - Y_{A_1}$ with $Y_{A_1}^\top Y_{A_2} = HQ$ a polar decomposition (known to be unique in the case when $Y_{A_1}^\top Y_{A_2}$ is nonsingular, which holds under Hypothesis 1). This curve has the following properties:

1. $\varphi_{A_1 \to A_2}(0) = A_1$, and $\varphi_{A_1 \to A_2}(1) = A_2$.
2. For each $t \in [0,1]$, $\varphi_{A_1 \to A_2}(t) \in \mathscr{S}_+(r, n)$; for $t \in \mathbb{R}\backslash[0,1]$, $\varphi_{A_1 \to A_2}(t) \in \mathscr{S}_+(\leq r, n)$,

where the notation $\mathscr{S}_+(\leq r, n)$ stands for the set of positive-semidefinite matrices of rank upper-bounded by r. We finally mention that [44] also contains expressions for the exponential and logarithm maps, which will allow us to implement the piecewise Bézier curves and surfaces used in this paper.

Instead of working directly on the quotient manifold (which involves the computation of geodesics, exponential and logarithm maps), a simpler approach consists in working on an affine section of the quotient, i.e., a submanifold of $\mathbb{R}_*^{n \times r}$ that intersects each orbit in at most one point, see [60]. Consider an equivalence class $[Y_A]$ with Y_A a representative of the class. We choose the section of the quotient at Y_A to be the set of points:

$$\mathscr{S}_A := \left\{ Y_A \left(I + (Y_A^\top Y_A)^{-1} S \right) + Y_{A\perp} K \mid S^\top = S, \, S \succ -Y_A^\top Y_A, \, K \in \mathbb{R}^{(n-r) \times r} \right\}, \tag{2.2}$$

where the matrix $Y_{A\perp} \in \mathbb{R}^{n \times (n-r)}$ is any orthonormal basis for the orthogonal complement of Y_A, i.e., $Y_A^\top Y_{A\perp} = 0$ and $Y_{A\perp}^\top Y_{A\perp} = I_{n-r}$. The constraint $S \succ -Y_A^\top Y_A$ guarantees that there is at most one representative of each equivalence class $[Y_B]$ in the section, and exactly one under the generic condition that $Y_A^T Y_B$ is nonsingular (that holds under Hypothesis 1).

Consider the section of the quotient at Y_{A_1}. The representative in the section of any equivalence class $[Y_{A_2}]$ (with $Y_{A_1}^T Y_{A_2}$ nonsingular) is then $\bar{Y}_{A_2} = Y_{A_2} Q^\top$, where Q is the orthogonal factor of the polar decomposition of $Y_{A_1}^\top Y_{A_2}$. Once all the points are projected on the section, we can simply interpolate these (projected) points using any Euclidean interpolation algorithm.

3 Construction of families of low-rank covariance matrices

The families of covariance matrices that we consider in this paper are the image of some covariance functions, i.e., multivariate functions defined by interpolation of a set of anchor points. The choice of these anchor points provides some flexibility in the design of the covariance families, allowing the latter to be tailored to the problem/dataset considered. More generally, we define a low-rank covariance function as a mapping from a set of parameters to a low-rank PSD matrix.

Definition 3.1 (Low-rank covariance function and family) A p-parameter low-rank covariance function is a map $\varphi : \mathbb{R}^p \to \cup_{k=0}^r \mathscr{S}_+(k,n)$, for $r < n$; its corresponding covariance family is the image of the covariance function φ.

Remark 3.1 The interpolation methods relying on a section (and not on geodesics) discussed in this paper may contain matrices of rank strictly lower than r as discussed in the next section.

3.1 First-order covariance functions

In this section, we consider two possible generalizations of multilinear interpolation to manifolds. The simplest way consists in mapping all the points to a linear approximation of the manifold (here, a section of the quotient), and applying multilinear interpolation on the section. A second approach resorts to the geodesics (generalization of straight lines) on the manifold. It is interesting to notice that both reduce to the interpolating geodesic in the one-parameter case if the reference point of the section belongs to the geodesic (cf. Remark 3.2 below).

3.1.1 First-order sectional covariance function

The main idea is to consider a section, to project the data matrices to that section as represented in Figure 3.1, and to perform multilinear interpolation on the section.

Definition 3.2 (The *sectional* p-parameter covariance function) Given data (i.e., anchor) matrices $A_1, \ldots, A_N$, with $N = 2^p$, the sectional p-parameter covariance function interpolating $A_1, \ldots, A_N$ at the corners of the hypercube $[0,1]^p$ is obtained as follows:

1. Compute $Y_1, \ldots, Y_N$, satisfying respectively $Y_1 Y_1^T = A_1, \ldots Y_N Y_N^T = A_N$.
2. Select a member $\bar{Y}_{A_1}$ of the equivalence class $[Y_{A_1}]$.
3. Compute $\bar{Y}_{A_i}, i = 2 \ldots N$, the intersection of the equivalence classes $[Y_i]$, $i = 2, \ldots, N$, with the section defined by $\bar{Y}_{A_1}$, as described in Section 2 (these points are well defined under Hypothesis 1).
4. Interpolate the matrices $\bar{Y}_{A_i}, i = 1 \ldots N$ (using classical Euclidean multilinear interpolation). For instance, in the bilinear case ($p = 2$ and $N = 4$):

$$\bar{Y}_{\varphi_{A_1 \to \cdots \to A_4}}(t_1, t_2) = \bar{Y}_{A_1}(1-t_1)(1-t_2) + \bar{Y}_{A_2}(1-t_1)t_2 + \bar{Y}_{A_3}t_1(1-t_2) + \bar{Y}_{A_4}t_1 t_2.$$

5. Generate (if required) the corresponding PSD matrix:

$$\varphi_{A_1 \to A_2 \to A_3 \to A_4}(t_1, t_2) = \bar{Y}_{\varphi_{A_1 \to A_2 \to A_3 \to A_4}}(t_1, t_2) \bar{Y}_{\varphi_{A_1 \to A_2 \to A_3 \to A_4}}^\top(t_1, t_2).$$
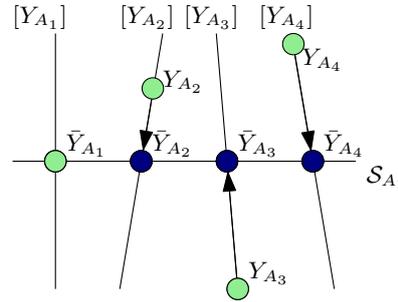
Fig. 3.1 Abstract representation of the selection of the equivalence classes representatives, when computing the sectional *p*-parameter covariance function. The vertical lines correspond to equivalence classes and the horizontal line is the section into which the data matrices are projected.

3.1.2 First-order geodesic covariance function

Instead of performing the interpolation on the section, we can also directly interpolate the points on the manifold, using geodesics. We define the *geodesic p-parameter covariance function*, represented in Figure 3.2.

Definition 3.3 (The *geodesic p*-parameter covariance function) Given data (i.e., anchor) matrices $A_1, \ldots, A_N$, with $N = 2^p$, the geodesic *p*-parameter covariance function interpolating interpolating $A_1, \ldots, A_N$ at the corners of the hypercube $[0,1]^p$ is obtained by combining geodesics between pairs of points (the latter are well defined under Hypothesis 1). For example, in the "bilinear" case ($p = 2$ and $N = 4$), the geodesic two-parameter covariance function is

$$\varphi_{(A_1 \to A_2) \to (A_3 \to A_4)}(t_1, t_2) := \varphi_{\left(\varphi_{A_1 \to A_2}(t_1)\right) \to \left(\varphi_{A_3 \to A_4}(t_1)\right)}(t_2). \tag{3.1}$$

Recursively, we can construct the geodesic *p*-parameter covariance function (in an analogous way as multilinear interpolation generalizes bilinear interpolation in the Euclidean space).
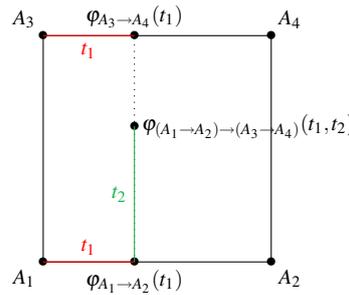
Fig. 3.2 The geodesic 2-parameter covariance function, in the specific case when the underlying manifold is a Euclidean space. For an arbitrary manifold, the construction process remains the same, straight lines becoming geodesics.

Remark 3.2 For the one-parameter case, if the reference point of the section is on the geodesic, the sectional and geodesic families coincide. Indeed, using the relationships from Section 2, notice that the geodesic (geodesic one-parameter function) can be converted to a sectional one-parameter covariance function:

$$Y_{\varphi_{A_1 \to A_2}}(t) = Y_{A_1} + t \dot{Y}_{A_1 \to A_2} = Y_{A_1} + t(Y_{A_2} Q^\top - Y_{A_1}) = Y_{A_1}(1 - t) + t Y_{A_2} Q^\top$$

$$= Y_{A_1}(1-t) + t\bar{Y}_{A_2},$$

where $\bar{Y}_{A_2} = Y_{A_2}Q^\top$ is the projection of Y_{A_2} into the section defined by Y_{A_1}.

3.2 Piecewise first-order covariance functions

The two p-dimensional interpolation schemes discussed above only use $N = 2^p$ data points. In applications, one often encounters situations where more data are available. This is for example the case of the wind field application discussed in Section 5: we assume the wind field in the area of interest to depend on two parameters (the prevailing wind orientation and magnitude), but we have more than four training points available. A typical way to take all these data into account is to use piecewise (if $p = 1$) or patchwise (if $p \geq 2$) interpolation. Let us consider the 2-dimensional case, without loss of generality. We assume that we have a collection of data (i.e., "anchor") matrices $(A_{i,j})$, with $A_{i,j} \in \mathscr{S}_+(r,n)$, and with $i \in \{0,\ldots,N_1\}$ and $j \in \{0,\ldots,N_2\}$. The two covariance families proposed in the previous section can be computed on each patch of the grid, to build patchwise multilinear surfaces. The resulting surfaces will be denoted φ^{LS} and φ^{LG}, where the L stands for Linear (those surfaces were obtained as generalization of linear interpolation on manifolds), the S for Section, and the G for Geodesic.

Definition 3.4 Let $(A_{i,j})$ be a data set, with $A_{i,j} \in \mathscr{S}_+(r,n)$, and with $i = 0,\ldots,N_1$, $j = 0,\ldots,N_2$. The first-order patchwise surfaces $\varphi^{\mathrm{LS}} : [0,N_1] \times [0,N_2] \to \mathscr{S}_+(\leq r,n)$ and $\varphi^{\mathrm{LG}} : [0,N_1] \times [0,N_2] \to \mathscr{S}_+(\leq r,n)$ are defined on each domain $[l,l+1] \times [m,m+1]$, with $l = 0,\ldots,N_1-1$ and $m = 0,\ldots,N_2-1$, as, respectively, the sectional two-parameter covariance function defined in Definition 3.2 and the geodesic two-parameter covariance function defined in Definition 3.3, interpolating the points $A_{l,m}$, $A_{l+1,m}$, $A_{l,m+1}$, $A_{l+1,m+1}$. When the surfaces computed on two adjacent patches do not coincide on a patch boundary, the value of the patchwise surface is determined by the patch with the largest indices (i.e., if the patch with indices $l = 0$, $m = 0$ is in the lower right corner of the grid, the value is given by the patch on the right/above the current patch).

Proposition 3.1 *There holds* $\varphi^{\mathrm{LS}}(t_1,t_2) \in \mathscr{S}_+(r,n)$ *and* $\varphi^{\mathrm{LG}}(t_1,t_2) \in \mathscr{S}_+(r,n)$.

Proof For φ^{LS}, this is a consequence of the fact that the multilinear Euclidean interpolation schemes only involve convex combinations of the data matrices, combined with the fact that the section (2.2) is a convex set. For φ^{LG}, this is a consequence of the fact that the interpolation scheme only involve geodesics, whose image belong to $\mathscr{S}_+(r,n)$ by definition, and that are well defined by Hypothesis 1.

3.3 Higher-order covariance functions: Bézier surfaces

The previous section presented two covariance families defined as generalizations of multilinear interpolation on the manifold. In this section, we consider higher-order interpolation on the manifold. While the methods in the previous sections were presented for an arbitrary number of parameters, we focus here on the two-parameter case ($p = 2$), as this is the situation encountered in the application of Section 5, and thus we resort to patchwise Bézier surfaces. Again, we will distinguish two cases: methods resorting to Euclidean interpolation in a section of the manifold and methods based on successive evaluations of geodesics.

3.3.1 Higher-order sectional covariance function

This method consists in first projecting all the data matrices on a section of the manifold (cf. Equation (2.2)), and then applying classical Euclidean Bézier interpolation in the section.

We focus on patchwise cubic Bézier surfaces. Each patch of those surfaces is a cubic Bézier surface defined according to a set of control points $(b_{i,j})_{i,j=0,\ldots,3}$, with $b_{i,j} \in \mathbb{R}^{n \times r}$ $\forall i,j$, as:

$$\beta : [0,1] \times [0,1] \to \mathbb{R}^{n \times r} :$$

$$(t_1,t_2) \mapsto \beta(t_1,t_2; (b_{i,j})_{i,j=0,\ldots,3}) := \sum_{i=0}^{3} \sum_{j=0}^{3} b_{i,j} B_{i,3}(t_1) B_{j,3}(t_2),$$

where $B_{i,3}(t) := \sum_{i=0}^{3} \binom{3}{i} t^i (1-t)^{3-i}$, with $t \in [0,1]$, is the Bernstein polynomial of order 3.

We define the patchwise cubic Bézier surface on the section, denoted φ^{BS}, as follows.

Definition 3.5 Let $(A_{i,j})$ be a data set, with $A_{i,j} \in \mathscr{S}_+(r,n)$, and with $i = 0,\ldots,N_1$, $j = 0,\ldots,N_2$. Choose a section of the quotient manifold and compute, for the equivalence class associated to each data matrix $A_{i,j}$, the representative $\bar{Y}_{i,j} \in \mathbb{R}_*^{n \times r}$ that belongs to the section, which is well-defined according to Hypothesis 1. Let $Y_{\varphi^{\mathrm{BS}}} : [0,N_1] \times [0,N_2] \to \mathbb{R}^{n \times r}$ be the patchwise Bézier surface interpolating the data matrices $(\bar{Y}_{i,j})$, and computed as in [2] (in which the manifold is simply the Euclidean space $\mathbb{R}^{n \times r}$). The surface φ^{BS} is obtained as follows.

$$\varphi^{\mathrm{BS}} : [0,N_1] \times [0,N_2] \to \mathscr{S}_+(\leq r,n) : (t_1,t_2) \mapsto Y_{\varphi^{\mathrm{BS}}}(t_1,t_2) Y_{\varphi^{\mathrm{BS}}}(t_1,t_2)^\top. \tag{3.2}$$

Proposition 3.2 *Apart from possibly a zero measure set, it holds that $\varphi^{\mathrm{BS}} \in \mathscr{S}_+(r,n)$.*

Proof Consider the function $f : [0,N_1] \times [0,N_2] \to \mathbb{R} : (t_1,t_2) \mapsto \det(Y_{\varphi^{\mathrm{BS}}}(t_1,t_2)^\top Y_{\varphi^{\mathrm{BS}}}(t_1,t_2))$. This function is zero if and only if $\varphi^{\mathrm{BS}}(t_1,t_2)$ has rank $k < r$. The result follows then from the fact that the restriction of $Y_{\varphi^{\mathrm{BS}}}(t_1,t_2)$ to each patch is a polynomial function of t_1 and t_2, which is real analytic.

3.3.2 Higher-order covariance function based on the exp and log

In this case, we refer to the Bézier surface interpolation algorithm on manifolds, proposed in [2]. This algorithm relies on the expressions for the logarithm and the exponential map, provided in [44] for the manifold $\mathscr{S}_+(r,n)$. We define the surface φ^{BG} (patchwise Bézier-like on the manifold) as follows.

Definition 3.6 Let $(A_{i,j})$ be a data set, with $A_{i,j} \in \mathscr{S}_+(r,n)$, and with $i = 0,\ldots,N_1$, $j = 0,\ldots,N_2$. The surface $\varphi^{\mathrm{BG}} : [0,N_1] \times [0,N_2] \to \mathscr{S}_+(\leq r,n)$ is the surface obtained by applying the Bézier surface interpolation algorithm on manifolds proposed in [2] (where the manifold is $\mathscr{S}_+(r,n)$ endowed with the Riemannian metric described in Section 2), with type-II reconstruction (cf. [2, Definition 4]), and with the control points chosen as suggested in [1]. This surface is well-defined according to Hypothesis 1.

Proposition 3.3 *There holds $\varphi^{\mathrm{BG}} \in \mathscr{S}_+(r,n)$.*

Proof This is a consequence of the fact that the function φ^{BG} is the image of an exponential map, that is well defined under Hypothesis 1.

3.4 Interpolation of labeled matrices using covariance functions

Note that all the covariance functions described here are defined by interpolation of a set of data matrices. Though the data matrices can be chosen by the user in any way, thus providing flexibility for problem-specific tailoring when the underlying stochastic process is parameter-dependent, they can in particular be chosen as the covariance matrices associated with a grid of parameter values. In that case, the elements of the covariance families are simply estimates of the covariance for other (intermediate) parameter values; i.e., they solve the following problem:

Problem 3.1 (Interpolation with low-rank covariance functions) Given an arbitrary set of $(N_1 + 1) \times (N_2 + 1)$ data matrices $A_{i,j} \in \mathscr{S}_+(r,n)$, associated with parameter values (i,j), for $i = 0, \ldots, N_1$ and $j = 0, \ldots, N_2$, what is the matrix corresponding to some parameter vector $x \in [0, N_1] \times [0, N_2]$?

We will show in the wind field application (see Section 5) that simply evaluating the covariance functions defined in this paper answers this problem with a reasonable accuracy.

4 Covariance identification using distance minimization

Having proposed multi-parameter low-rank covariance families in the previous section, we can now describe identification procedures within such families. That is, given a data set $(y_i)_{i=1}^q$ (assumed to be centered, with $y_i \in \mathbb{R}^n$) from which we construct a sample covariance matrix $\widehat{C} = \frac{1}{q} \sum_i y_i y_i^\top$, we would like to find the most representative member of a family.

A widely used methodology for selecting a representative member of a covariance family is maximum likelihood estimation. Let $\varphi_{A_1,\ldots,A_m}(t)$ denote the covariance family being considered. In maximum likelihood estimation, the data are assumed to have a certain probability distribution, e.g., $y_i \sim N\big(0, \varphi_{A_1,\ldots,A_m}(t)\big)$ for all i, and we choose $t \in \mathbb{R}^p$ to maximize the resulting likelihood function $p(y_1, \ldots, y_q | t)$. Since the covariance matrices $\varphi_{A_1,\ldots,A_m}(t)$, for each t, are low-rank, the Gaussian distribution of the data is singular. Thus the associated probability density function is not well-defined for generic y_i and maximizing the likelihood becomes non-trivial. (If the matrices were instead SPD and as $q \to \infty$, this problem is equivalent to minimizing the Kullback–Leibler divergence of $N(0,\widehat{C})$ from $N\big(0, \varphi_{A_1,\ldots,A_m}(t)\big)$, known as reverse information projection [20, 10, 15].)

Rather than maximizing the likelihood, we propose to minimize the Frobenius distance between the covariance function and the sample covariance matrix $\widehat{C}$.

Problem 4.1 (Minimum Frobenius distance covariance identification)

$$\arg\min_{t \in \mathbb{R}^p} d_F\big(\varphi_{A_1,\ldots,A_m}(t), \widehat{C}\big),$$

where $d_F(A_1, A_2) = \|A_1 - A_2\|_F$ is the Frobenius distance.

This is a particular instance of "minimum distance estimation"; such estimators, in general, have a long history in statistics [66]. Our main motivations for choosing the Frobenius distance, in contrast with, e.g., the distance naturally inherited from the Riemannian metric considered on $\mathscr{S}_+(r,n)$, are the resulting simplicity of the objective function to minimize, and its applicability to the small data regime. Note that Problem 4.1 does not make assumptions on the relative values of q and r, and in particular does not require q to be larger than r (or thus n).

We now discuss solutions to Problem 4.1 given particular constructions of the covariance function $\varphi(t)$. Since the geodesic and sectional approaches to define covariance functions coincide for the one-parameter case under some conditions (cf. Remark 3.2), we divide this section in two parts: the one-parameter case ($p = 1$) and the multi-parameter case. For the latter, we focus on the case of two parameters ($p = 2$).

4.1 One-parameter first-order covariance function

For $p = 1$, the covariance function $t_1 \mapsto \varphi(t_1)$ is simply the geodesic between the two data matrices. The optimization problem has a closed form solution that is presented below.

Proposition 4.1 (Solution of the low-rank covariance identification problem) *The solutions of Problem 4.1 for $p = 1$ are the roots of a third order polynomial $at^3 + bt^2 + ct + d = 0$ with:*

$$a = 4\mathrm{tr}\left(\dot{Y}_{A_1 \to A_2} \dot{Y}_{A_1 \to A_2}^\top \dot{Y}_{A_1 \to A_2} \dot{Y}_{A_1 \to A_2}^\top \right),$$

$$b = 12\mathrm{tr}\left(Y_{A_1} \dot{Y}_{A_1 \to A_2}^\top \dot{Y}_{A_1 \to A_2} \dot{Y}_{A_1 \to A_2}^\top \right),$$

$$c = 4\mathrm{tr}\left(2Y_{A_1} Y_{A_1}^\top \dot{Y}_{A_1 \to A_2} \dot{Y}_{A_1 \to A_2}^\top + Y_{A_1} \dot{Y}_{A_1 \to A_2}^\top Y_{A_1} \dot{Y}_{A_1 \to A_2}^\top - \dot{Y}_{A_1 \to A_2} \dot{Y}_{A_1 \to A_2}^\top Y_{\widehat{C}} Y_{\widehat{C}}^\top \right),$$

$$d = 4\mathrm{tr}\left(Y_{A_1} Y_{A_1}^\top \dot{Y}_{A_1 \to A_2} Y_{A_1}^\top - \dot{Y}_{A_1 \to A_2} Y_{A_1}^\top Y_{\widehat{C}} Y_{\widehat{C}}^\top \right).$$

Proof The cost function is:

$$d_F\left(\varphi_{A_1 \to A_2}(t), \widehat{C} \right) = \sqrt{\mathrm{tr}\left(\left(\varphi_{A_1 \to A_2}(t) - \widehat{C} \right)\left(\varphi_{A_1 \to A_2}(t) - \widehat{C} \right)^\top \right)}.$$

The third-order polynomial is obtained after setting the derivative to zero and noting that the optimization problem is unconstrained.

Proposition 4.1 provides at least one solution. If there are three roots, the minimizer is of course the one with smallest objective. As with any third order polynomial, the uniqueness condition of this solution is:

$$18abcd - 4b^3d + b^2c^2 - 4ac^3 - 27a^2d^2 \leq 0.$$

The computational cost of finding the solution of the low-rank covariance identification problem is only $\mathcal{O}(nr)$. Indeed, roots of the cubic equation have a closed form expression whose evaluation does not require any meaningful cost. The only computational cost is that associated with computing traces to obtain the polynomial coefficients. By virtue of the cyclic property of the trace, we can compute these traces with $\mathcal{O}(nr)$ elementary operations.

4.2 Two-parameter first-order covariance functions

Here we focus on Problem 4.1 in the two-parameter case ($p = 2$) for first-order covariance functions, i.e., $\varphi = \varphi^{\mathrm{LS}}$ or $\varphi = \varphi^{\mathrm{LG}}$ (cf. Definition 3.4).

Similarly to the previous sections, we assume that data matrices are defined on a grid of points $(A_{i,j})$, with $A_{i,j} \in \mathscr{S}_+(r,n,)$, $i = 0, \ldots, N_1$ and $j = 0, \ldots, N_2$.

4.2.1 First-order sectional covariance function

In the case $\varphi = \varphi^{\mathrm{LS}}$ (see Definition 3.4), we propose to use a gradient descent on each patch of the surface. Observe indeed that the surface φ^{LS} is generally nondifferentiable (actually, even noncontinuous) on the borders of the patches. The global optimum is then computed as the minimum of the minima obtained on the patches. Let

$$f : [0, N_1] \times [0, N_2] \to \mathbb{R} : (t_1, t_2) \mapsto f(t_1, t_2) := d_F(\varphi^{\mathrm{LS}}(t_1, t_2), \widehat{C})^2 \tag{4.1}$$

be the cost function to minimize.

Consider an arbitrary patch (l, m), with $l = 0, \ldots, N_1 - 1$ and $m = 0, \ldots, N_2 - 1$. Let $f^{l,m}$ be the restriction of f to the patch (l, m), and let $A_{l,m}, A_{l+1,m}, A_{l,m+1}, A_{l+1,m+1}$ denote the four corners of the patch (l, m), and $\bar{Y}_{A_{l,m}}, \bar{Y}_{A_{l+1,m}}, \bar{Y}_{A_{l,m+1}}, \bar{Y}_{A_{l+1,m+1}}$ their projection on the section. We omit the superscript LS in

the remainder of this section. The gradient of the restriction of the cost function (4.1) to that patch can be computed explicitly:

$$\frac{\partial f^{l,m}}{\partial t_1}(t_1,t_2) = 2\mathrm{tr}\left(\frac{\partial \varphi}{\partial t_1}(t_1,t_2)(\varphi(t_1,t_2)-\widehat{C})^\top\right), \tag{4.2}$$

$$\frac{\partial f^{l,m}}{\partial t_2}(t_1,t_2) = 2\mathrm{tr}\left(\frac{\partial \varphi}{\partial t_2}(t_1,t_2)(\varphi(t_1,t_2)-\widehat{C})^\top\right), \tag{4.3}$$

with $\frac{\partial \varphi}{\partial t_1}(t_1,t_2)$ and $\frac{\partial \varphi}{\partial t_2}(t_1,t_2)$ computed as:

$$\frac{\partial \varphi}{\partial t_1}(t_1,t_2) = \frac{\partial Y_\varphi}{\partial t_1}(t_1,t_2)Y_\varphi^\top + Y_\varphi\left(\frac{\partial Y_\varphi}{\partial t_1}(t_1,t_2)\right)^\top,$$

$$\frac{\partial \varphi}{\partial t_2}(t_1,t_2) = \frac{\partial Y_\varphi}{\partial t_2}(t_1,t_2)Y_\varphi^\top + Y_\varphi\left(\frac{\partial Y_\varphi}{\partial t_2}(t_1,t_2)\right)^\top.$$

The factor Y_φ was defined in Definition 3.2:

$$Y_\varphi = \bar{Y}_{A_{l,m}}(1-t_1)(1-t_2) + \bar{Y}_{A_{l+1,m}}(1-t_1)t_2 + \bar{Y}_{A_{l,m+1}}t_1(1-t_2) + \bar{Y}_{A_{l+1,m+1}}t_1t_2.$$

So $\frac{\partial Y_\varphi}{\partial t_1}(t_1,t_2)$ and $\frac{\partial Y_\varphi}{\partial t_2}(t_1,t_2)$ can be easily obtained as:

$$\frac{\partial Y_\varphi}{\partial t_1}(t_1,t_2) = t_2(\bar{Y}_{A_{l,m}} - \bar{Y}_{A_{l+1,m}} - \bar{Y}_{A_{l,m+1}} + \bar{Y}_{A_{l+1,m+1}}) + \bar{Y}_{A_{l,m+1}} - \bar{Y}_{A_{l,m}},$$

$$\frac{\partial Y_\varphi}{\partial t_2}(t_1,t_2) = t_1(\bar{Y}_{A_{l,m}} - \bar{Y}_{A_{l+1,m}} - \bar{Y}_{A_{l,m+1}} + \bar{Y}_{A_{l+1,m+1}}) + \bar{Y}_{A_{l+1,m}} - \bar{Y}_{A_{l,m}},$$

with the only difference being the parameter t_1 vs. t_2.

4.2.2 First-order geodesic covariance function

We focus now on Problem 4.1 for $p = 2$, when the covariance function is the surface φ^{LG} defined in Definition 3.4. Let

$$f : [0,N_1] \times [0,N_2] \to \mathbb{R} : (t_1,t_2) \mapsto f(t_1,t_2) := d_F(\varphi^{\mathrm{LG}}(t_1,t_2),\widehat{C})^2 \tag{4.4}$$

be the cost function. The surface φ^{LG} is generally not differentiable on the borders of the patches. As a result, similarly to the previous section, we propose to run an optimization algorithm to find the optimum on each patch, and to compare the optimal values obtained on the patches to obtain the global optimum.

Let $f^{l,m}$ be the restriction of f to the patch (l,m). We propose to minimize $f^{l,m}$ by expressing it as a one-variable function, replacing t_2 by its optimal value:

$$t_2^*(t_1) = \operatorname*{argmin}_{t_2 \in \mathbb{R}} f^{l,m}(t_1,t_2), \tag{4.5}$$

and then to apply gradient descent to the problem:

$$\min_{t_1 \in \mathbb{R}} \tilde{f}^{l,m}(t_1) := f^{l,m}(t_1,t_2^*(t_1)). \tag{4.6}$$

The computation of the partial derivatives required by both steps is deferred to Appendix A.

4.3 Higher-order covariance functions using Bézier curves

We now solve Problem 4.1 for $p = 2$ when the surface is defined from Bézier interpolating surfaces.

4.3.1 Higher-order sectional covariance function

To solve Problem 4.1 for $p = 2$ when the surface is a Euclidean Bézier surface built in a given section of the manifold, we propose again to use steepest descent. The cost function

$$f : [0, N_1] \times [0, N_2] \to \mathbb{R} : (t_1, t_2) \mapsto f(t_1, t_2) := d_F(\varphi^{\mathrm{BS}}(t_1, t_2), \widehat{C})^2, \tag{4.7}$$

with φ^{BS} defined in Definition 3.5, is $\mathscr{C}^1$. Moreover, since Bézier curves in the Euclidean space are weighted sums of Bernstein polynomials, the gradient can be computed explicitly. The computation of the gradient is similar to Section 4.2.1, except that now Y_φ is obtained as a linear combination of cubic Bernstein polynomials:

$$Y_\varphi(t_1, t_2) = \sum_{i=0}^{3} \sum_{j=0}^{3} b_{ij} B_{i3}(t_1) B_{j3}(t_2).$$

The derivatives $\frac{\partial Y_\varphi}{\partial t_1}(t_1, t_2)$ and $\frac{\partial Y_\varphi}{\partial t_1}(t_1, t_2)$ become:

$$\frac{\partial Y_\varphi}{\partial t_1}(t_1, t_2) = \sum_{i=0}^{3} \sum_{j=0}^{3} b_{ij} \dot{B}_{i3}(t_1) B_{j3}(t_2),$$

$$\frac{\partial Y_\varphi}{\partial t_2}(t_1, t_2) = \sum_{i=0}^{3} \sum_{j=0}^{3} b_{ij} B_{i3}(t_1) \dot{B}_{j3}(t_2).$$

4.3.2 Higher-order covariance function based on the exponential and logarithm maps

For φ^{BG} in Definition 3.6, it remains unclear whether the gradient of the cost function:

$$f : [0, N_1] \times [0, N_2] \to \mathbb{R} : (t_1, t_2) \mapsto d_F(\varphi^{\mathrm{BG}}(t_1, t_2), \widehat{C})^2. \tag{4.8}$$

has an analytical expression. Variable projection methods also do not seem applicable in this case. Thus we have to estimate the gradient numerically, resorting to finite differences.

5 Case study: wind field approximation

Given the increasing popularity of unmanned aerial vehicles (UAVs) in transportation, surveillance, agriculture, and beyond, accurate and safe aerial navigation is essential. Achieving these requirements demands expressive models of the UAV's environment—in particular, the wind field—and the ability to update these models given new observations, e.g., via Kalman filtering [51, 16]. To this end, we wish to construct and estimate the covariance of spatially distributed wind velocity components.

5.1 Model problem and data set

Gaussian random field models have previously been used to describe spatially varying wind velocities ([67, 35]), in addition to the other applications mentioned in the introduction (e.g, facial recognition [55], computer vision [40]). A common practice in this setting is to define the covariance matrix of the velocities using some standard covariance kernel, e.g., the squared-exponential kernel, perhaps with some modifications to allow for non-stationarity [36]. We instead assume to have instances of the covariance matrix for different values of the prevailing wind heading (angle of incidence) θ and magnitude (norm of the velocity vector) W; from these instances, we will build a covariance family for continuous (θ, W). The wind field can change dramatically as function of the prevailing wind, and thus it is useful to consider a covariance family built from a variety of representative prevailing wind settings.

In general, these instances could be estimated from observational data, or they could be constructed using offline (and potentially expensive) computational fluid dynamics simulations. Here we use the latter: we solve the unsteady incompressible Navier–Stokes equations on the two-dimensional domain shown in Figure 5.1, using direct numerical simulation with a spectral element method [30]. The Reynolds number in our simulations, defined according to the side-length of the central obstacle, is around 500 for $W = 7.0$. For each chosen value of (θ, W), we run the simulation until any transients due to the initial condition have dissipated and then collect 1000 instantaneous velocity fields as "snapshots," with examples shown in Figure 5.2. The sample covariance of these snapshots provides the data covariance matrix at that (θ, W).

The right plot of Figure 5.1 represents a notional idea of our example domain: flow around a rectangular cuboid in three dimensions. We consider only a horizontal "slice" of this domain, e.g., the wind in a plane at height h sufficiently far from the ground and from top of the obstacle so that a two-dimensional approximation is reasonable. The left plot of Figure 5.1 shows the mean value of the velocity on this plane, at an example value of (θ, W). The grid size is 39×39, and hence the discretized wind field has $n = 3024 = 2 \times (39^2 - 9)$ degrees of freedom: two velocity components at each grid point, subtracting 9 points for the obstacle. The grey contours represent the pointwise variance of the x-velocity plus that of the y-velocity (i.e., the sum of two diagonal entries of the covariance matrix, at each point in space). Naturally, the variance is larger downstream of the obstacle, where vortices are shed.

Our data set for the examples below comprises a set of covariance matrices $C(\theta_k, W_i)$, with $\theta_k = (k-1)\pi/64$, $k \in 1, \ldots, 32$ and $W \in \{4.0, 5.5, 7.0, 8.5, 10.0, 11.5, 13.0\}$, as illustrated in Figure 5.6. Using a truncated singular value decomposition of each matrix, we reduce the rank to $r = 20$. These covariance matrices then belong to $\mathscr{S}_+(20, 3024)$.

5.2 One-parameter covariance families

We first consider interpolation and identification with a one-parameter geodesic covariance function $\varphi_{A_1 \to A_2}(t)$, where the data matrices A_1 and A_2 are obtained at the same wind magnitude but at faraway headings: $A_1 = C(\theta_1 = 0, W = 8.5)$ and $A_2 = C(\theta_9 = 23, W = 8.5)$. (As noted in Remark 3.2, the one-parameter geodesic and sectional families coincide.) To understand the relationship between the geodesic parameter t and the true wind heading, we minimize the distance from each of the seven intermediate data matrices $C(\theta_k, 8.5)$, $k \in 1, 2, \ldots, 9$, to this covariance family (cf. red line in Figure 5.6) and obtain a value of t_k. Figure 5.3 shows the resulting pairs (θ_k, t_k). The relationship between t and θ is monotone and nearly linear. Similar results can be obtained for other choices of W.

Next we focus on the shape of the objective function used in distance minimization for one-parameter covariance families. We build a one-parameter covariance function $\varphi_{A_1 \to A_2}(t)$ with $A_1 = C(16.9, 7)$ and $A_2 = C(22.5, 7)$ and evaluate the distance to $A_3 = C(19.7, 7)$ as a function of $t \in [0, 1]$. (See Figure 5.6, dashed blue line, to identify the relevant matrices in our data set.) This exercise is shown in Figure 5.4, where the anchor or data matrices A_1, A_2 are illustrated via inset plots with a green obstacle. (The matrices
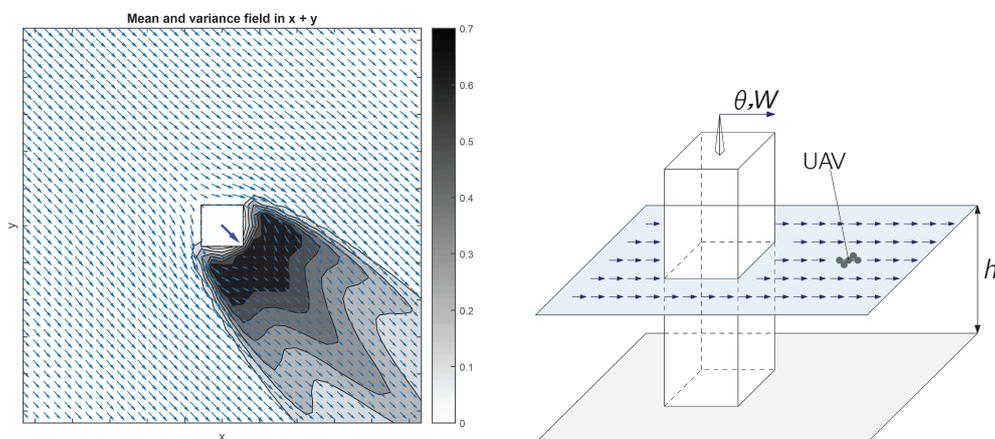
Fig. 5.1 Representation of the wind field. Left: two-dimensional domain, with wind field around the square obstacle represented by light blue arrows and the prevailing wind in dark blue; gray contours are the variance field. Right: notional 3-D problem, with a section of the wind field at an altitude h.

are visualized by their variance fields, as in Figure 5.1 (left).) First, we note that the distance objective is smooth and convex (on $[0,1]$), and that its minimum (marked with a blue dot) is close to, though not exactly, $t = 0.5$. This offset is a further instance of the difference illustrated in Figure 5.3, between the minimum-distance points and a perfect linear relationship. The inset plots in Figure 5.4 with white obstacles show covariances in the one-parameter family at intermediate values of t; we see that these covariances look physically reasonable, suggesting intermediate wind headings as desired. Nonetheless, we also note that the minimum Frobenius distance from A_3 to this family is roughly 14, about half of the distance from A_3 to the anchor A_1. For a more accurate representation of A_3, one may thus want a richer family or more representative data matrices. We will explore these choices below.

5.3 Distances to two-parameter covariance families

Now we illustrate the distance between a given matrix and two different two-parameter covariance families, each constructed from the same four data matrices. The minimizer of this distance is a solution of Problem 4.1.

First, we consider the first-order sectional covariance function of Section 3.1. We use four data matrices: $A_1 = C(11.3, 4)$, $A_2 = C(11.3, 10)$, $A_3 = C(16.9, 4)$, $A_4 = C(16.9, 10)$. Figure 5.5 (left) illustrates the distance from $\varphi^{\mathrm{LS}}_{(A_1 \to A_2) \to (A_3 \to A_4)}(t_1, t_2)$ to $A_5 = C(14.1, 7)$. The red triangle represents distance minimizer, which lies at $t_1 = 0.48$ and $t_2 = 0.77$ and yields a distance of 5.5 from the target. To define the section in this case, we use the matrix A_1. The four anchor matrices are the edges of the rectangle in Figure 5.6; other choices would lead to similar results, as analyzed in subsequent subsections.

Next, we repeat the study for the geodesic two-parameter covariance function defined in Section 3.1.2, with results shown in Figure 5.5 (right). Again, the minimizer is marked with a red triangle, which lies at $t_1 = 0.77$ and $t_2 = 0.48$ and yields a distance of 5.5 from the target. Note that the inputs to both covariance functions can in principle be any element of $\mathbb{R}^2$; here, both figures show the distance for $(t_1, t_2) \in [0,2] \times [-1,1]$. The distance contours are shaped slightly differently for the sectional and geodesic cases, though the minimizer lies in the top left quadrant $[0,1]^2$ of each figure, as expected.
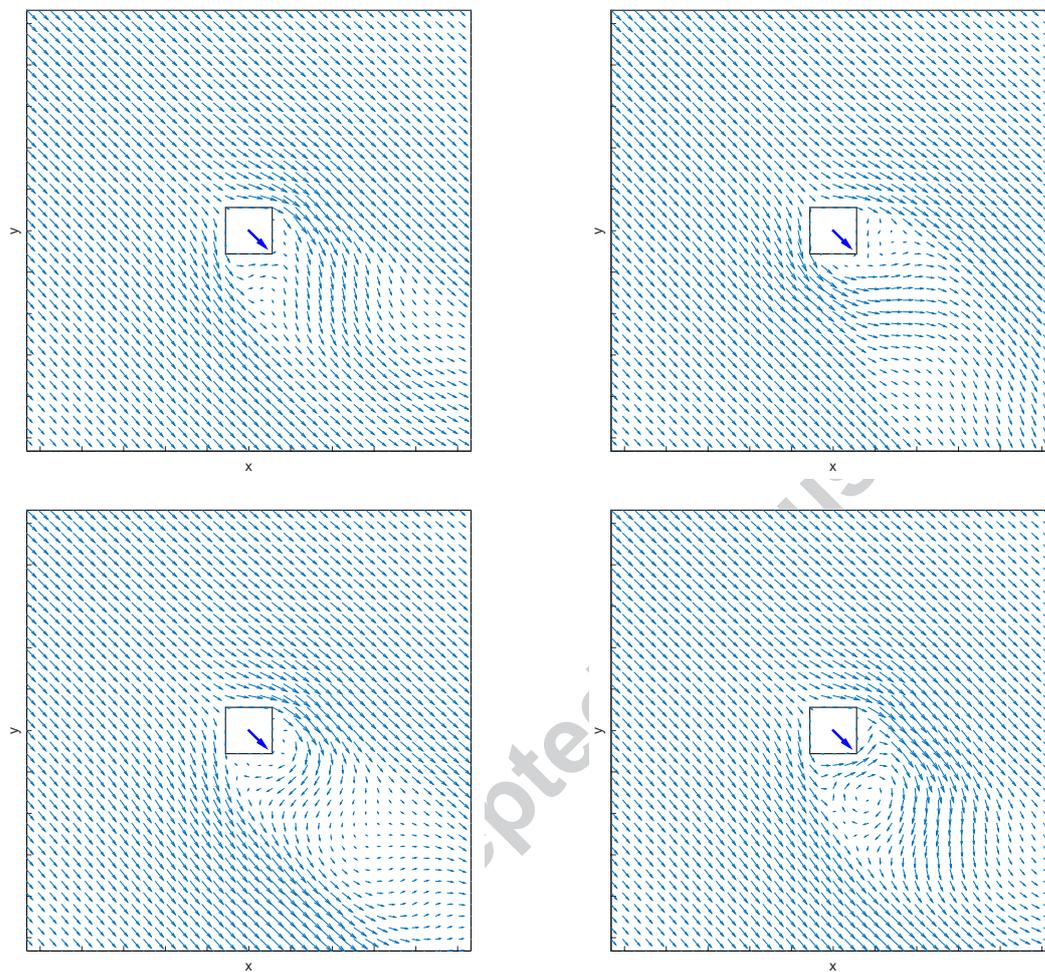
**Fig. 5.2** Instantaneous snapshots of the wind velocity field for $\theta = 45$ and $W = 7$.

5.4 Benchmarking first-order and higher-order covariance functions

Now we consider the four surfaces defined in Section 3: the first-order covariance functions φ^{LS} and φ^{LG} defined patchwise (see Definition 3.4), and the Bézier-like covariance functions φ^{BS} and φ^{BG} (see Definitions 3.5 and 3.6).

For the surfaces defined on a section of the manifold, we consider several possibilities: for φ^{LS}, the section based at one of the data matrices (here, the lower left data matrix of the patch), based at the arithmetic mean of the data matrices, or based at the inductive mean of the four data matrices of the patch. The latter, well-known in the literature on positive-definite matrices (see, e.g., [44] and references therein), is constructed recursively, using successive evaluations of geodesics:

$$M_{Ind} := \left(\left((A_1 \#_{1/2} A_2) \#_{1/3} A_3 \right) \dots \right) \#_{1/N} A_N, \tag{5.1}$$

where the notation $A_1 \#_t A_2$ is simply a shorthand for the geodesic $\varphi_{A_1 \to A_2}(t)$; see (2.1). Note that, in the Euclidean setting, geodesics become straight lines and the inductive mean coincides with the arithmetic mean. On (non-flat) manifolds, the inductive mean loses interesting properties such as invariance under
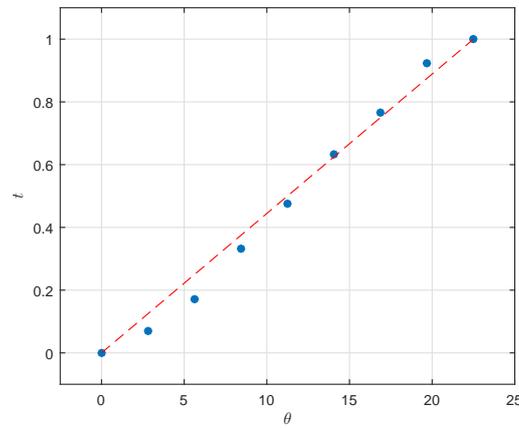
Fig. 5.3 Minimizing value of t (blue points) for data drawn from a range of wind headings θ, for $W = 8.5$ (cf. Section 5.2). The red line represents a "perfect" linear relationship.
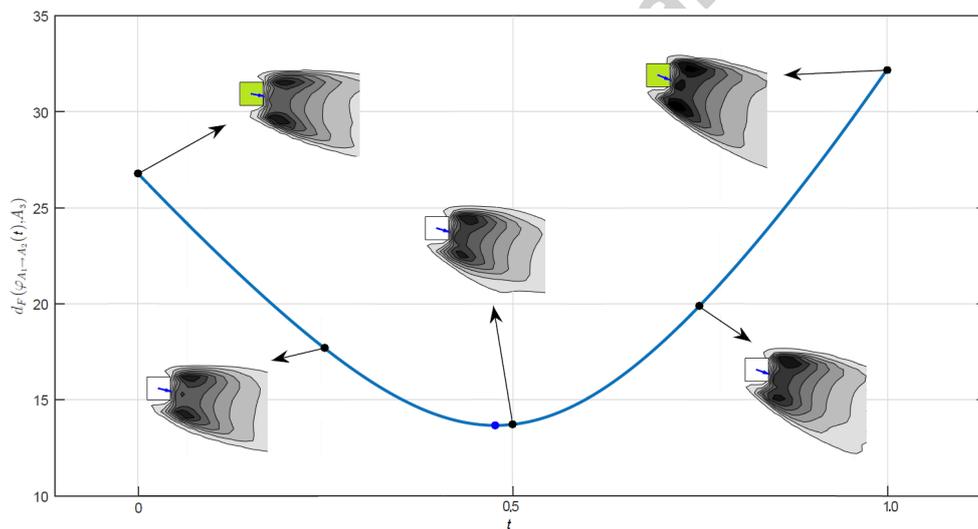

Fig. 5.4 Distance from $A_3 = C(19.7, 7)$ to the one-parameter covariance family built from $A_1 = C(16.9, 7)$ and $A_2 = C(22.5, 7)$, as a function of the input variable t. A_1 and A_2 are marked with green obstacles to identify them as the data matrices/anchors. The blue point represents the distance minimizer.

permutations of the data matrices. There exist other notions of means on the manifold of fixed-rank positive-semidefinite matrices, such as the Riemannian barycenter, but the computation of the latter is computationally more involved and not considered here. For φ^{BS}, the section is based at one of the data matrices (here, the lower left data matrix of the training set), at the arithmetic mean of the data matrices of the training set, or at the inductive mean of the four data matrices of the training set.

These combinations lead to a total of eight surfaces. The data matrices are split into two sets shown in Figure 5.6: the blue points and the black points. The blue points are used to construct the surface, and the accuracy of the methods is evaluated on the black points.
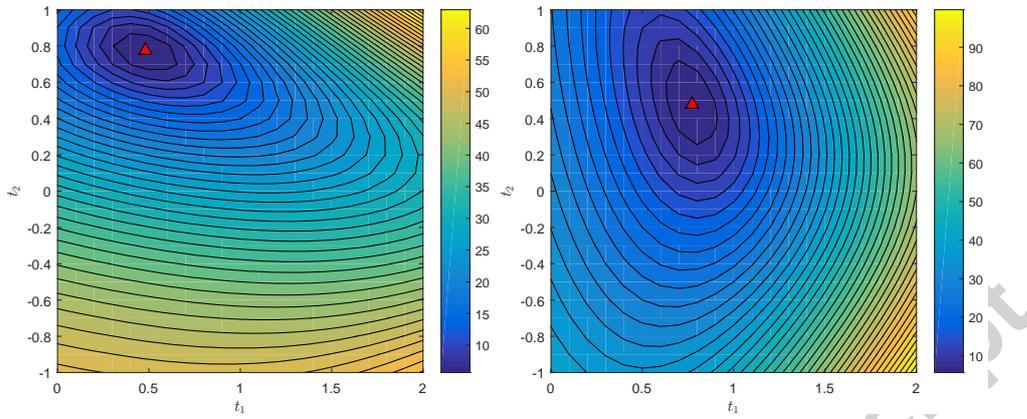
Fig. 5.5 Left: Distance from A_5 to $\varphi^{\text{LS}}_{(A_1 \to A_2) \to (A_3 \to A_4)}(t_1, t_2)$. Right: Distance from A_5 to $\varphi^{\text{LG}}_{(A_1 \to A_2) \to (A_3 \to A_4)}(t_1, t_2)$.
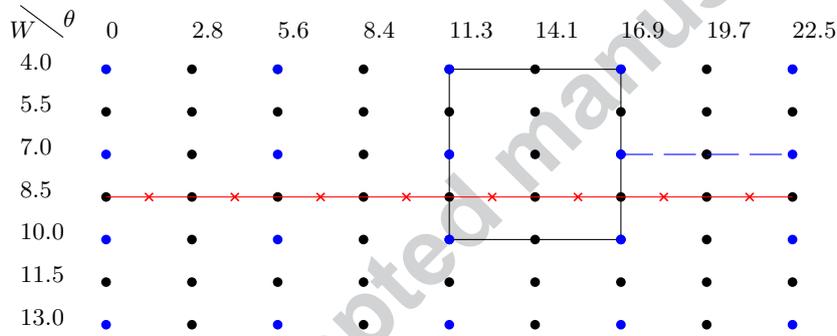
Fig. 5.6 Data set for the wind field problem; each dot shows the wind magnitude W and heading θ corresponding to a data covariance matrix. The red (crossed) and blue (dashed) lines represent the data used in Section 5.2. The rectangle illustrates the operating zone of Section 5.3. The blue nodes comprise the training set and the black nodes the test set for Sections 5.4.1 and 5.4.2.

5.4.1 Interpolation errors

The error

$$E\left(C(\theta, W)\right) := \left\| C(\theta, W) - \varphi^{\text{method}}(\theta, W) \right\|_F^2$$

is a measure of the ability of the surface φ^{method} to recover some hidden covariance matrix $C(\theta, W)$.[1] We will also consider the normalized error (in percentages)

$$E_N\left(C(\theta, W)\right) := 100 \times \frac{\left\| C(\theta, W) - \varphi^{\text{method}}(\theta, W) \right\|_F^2}{\frac{1}{4} \sum_{j=1}^4 \left\| C(\theta, W) - A_j \right\|_F^2},$$

where normalization is performed with respect to the average squared distance between the target matrix $C(\theta, W)$ and the four corners of the patch to which it belongs, according to the grid representation in Figure 5.6. (The patch is chosen systematically as the one below and to the left of the test point considered.)

Having defined these interpolation errors for arbitrary $C(\theta, W)$, we now evaluate them for all the points $C(\theta_i, W_i)$ in the test set—i.e., for each of the data matrices marked with black nodes in Figure 5.6. We average the errors over the test set and report the resulting values in Table 5.1.

---

[1] Here we write φ^{method} with arguments (θ, W) in a slight abuse of notation. More precisely, we mean that the two-parameter covariance function φ^{method} is evaluated at (t_1, t_2) corresponding to an affine mapping from the range of (θ, W) (here $[0, 22.5] \times [4, 13]$) to $[0, 4] \times [0, 3]$, consistent with the 5×4 grid of data matrices.

Table 5.1 Average (squared) distance separating a given test point $C(\theta_i, W_i)$ from the corresponding interpolation point on the different surfaces. For the methods defined on a section of the manifold, the subscript of φ indicates how the section was chosen: 'one' means that we use one of the data matrices (here, the matrix at the lower left corner of the patch) as basis of the section, while 'arith' and 'inductive' denote, respectively, the arithmetic and inductive means of the corners of the patch.

	$\mathrm{avg}_i\left[E\left(C(\theta_i, W_i)\right)\right]$	$\mathrm{avg}_i\left[E_N\left(C(\theta_i, W_i)\right)\right]$
1st-order section $\varphi_{\mathrm{one}}^{\mathrm{LS}}$	30.3	7.78
1st-order section $\varphi_{\mathrm{arithm}}^{\mathrm{LS}}$	30.4	7.80
1st-order section $\varphi_{\mathrm{inductive}}^{\mathrm{LS}}$	30.3	7.77
1st-order geodesic φ^{LG}	30.3	7.77
Bézier section $\varphi_{\mathrm{one}}^{\mathrm{BS}}$	20.8	4.91
Bézier section $\varphi_{\mathrm{arithm}}^{\mathrm{BS}}$	20.5	4.87
Bézier section $\varphi_{\mathrm{inductive}}^{\mathrm{BS}}$	20.4	4.85
Bézier geodesic φ^{BG}	20.3	4.79

Some takeaways from this study are as follows. First, the matrix chosen to define the section in either the first-order sectional covariance function or the higher-order sectional covariance function seems to have little impact. Moreover, the performance of the geodesic covariance functions is not noticeably better than that of the sectional functions in this setting. But the interpolation performance of the higher-order (Bézier) families is significantly better than that of the first-order.

Based on our experience, normalized errors of less than 10% are impactful for practical applications. This level means that we can achieve $10\times$ better accuracy (defined in terms of Frobenius distance) by using the interpolated matrices rather than the nearest anchors. With Bézier families, the error is less than 5%, which implies $20\times$ improved performance.

In terms of computational costs, higher-order interpolation is more intense than first-order. However, most of the computations for the Bézier curve can be performed offline. In fact, the reconstruction step in [2] is the only one that needs to be done online.

5.4.2 Identification errors and data compression

We now assess identification errors within the covariance families. In other words, we now use the techniques of Section 4 to minimize the distance from each element of the test set to the covariance family $\varphi^{\mathrm{method}}$, built patchwise from the training matrices. From another perspective, this process can be viewed as data compression: a simple way to perform data compression consists of storing only several matrices (in our case, the training data) and then storing, for any additional matrix, the coordinates of the closest point in the surface. We now compare our surfaces for this task. Similarly to the previous section, we use the following two error measures,

$$E^*\left(C(\theta_i, W_i)\right) = \left\| C(\theta_i, W_i) - \varphi(t_1^*(i), t_2^*(i)) \right\|_F^2,$$

$$E_N^*\left(C(\theta_i, W_i)\right) = \frac{100 \times \left\| C(\theta_i, W_i) - \varphi(t_1^*(i), t_2^*(i)) \right\|_F^2}{\frac{1}{4}\sum_{j=1}^4 \left\| C(\theta_i, W_i) - A_j \right\|_F^2},$$

where A_j are the four corners of the patch to which $C(\theta_i, W_i)$ belongs and $t_1^*(i), t_2^*(i)$ are the solutions to the optimization problem discussed in Section 4.

We evaluate these errors for every test matrix and report, in Table 5.2, the average errors for each surface definition proposed. These are essentially the average distances between an element of our test set and the closest point on the surface. A key takeaway from this table is that the errors are significantly lower than those in Table 5.1; this is not surprising, as here we are optimizing to find the best point in each family. Also, results with the geodesic families in this example appear to be slightly better than with the sectional families.

Table 5.2 Average (squared) distance separating a given test point from its closest approximation on the different surfaces. For methods defined on a section of the manifold, the subscript of φ indicates how the section was chosen: 'one' means that we use one of the data matrices (here, the matrix at the lower left corner of the patch) as basis of the section, while 'arithm' and 'inductive' denote, respectively, the arithmetic and inductive means of the corners of the patch.

	$\mathrm{avg}_i\left[E^*(C(\theta_i, W_i))\right]$	$\mathrm{avg}_i\left[E_N^*(C(\theta_i, W_i))\right]$
1st-order section $\varphi_{\mathrm{one}}^{\mathrm{LS}}$	21.8	5.30
1st-order section $\varphi_{\mathrm{arithm}}^{\mathrm{LS}}$	21.8	5.29
1st-order section $\varphi_{\mathrm{inductive}}^{\mathrm{LS}}$	21.8	5.30
1st-order geodesic φ^{LG}	20.9	5.10
Bézier section $\varphi_{\mathrm{one}}^{\mathrm{BS}}$	14.3	3.34
Bézier section $\varphi_{\mathrm{arithm}}^{\mathrm{BS}}$	14.0	3.29
Bézier section $\varphi_{\mathrm{inductive}}^{\mathrm{BS}}$	14.0	3.29
Bézier geodesic φ^{BG}	13.8	3.24

It is instructive to see how the normalized errors $E_N^*\left(C(\theta_i, W_i)\right)$ are distributed over the data set, i.e., how the approximation error depends on the parameters of the data matrices $C(\theta_i, W_i)$. We illustrate this distribution using the stem plot in Figure 5.7, for φ^{BG} only. Errors are largest for wind field headings between those of the training set (i.e., $\theta \in \{2.8, 8.4, 14.1, 19.7\}$) and increase strongly with the wind field magnitude. These trends indicate that it might be useful to generate a denser grid of data matrices in the θ direction, particularly for large W, and to use a coarser grid in the W direction.
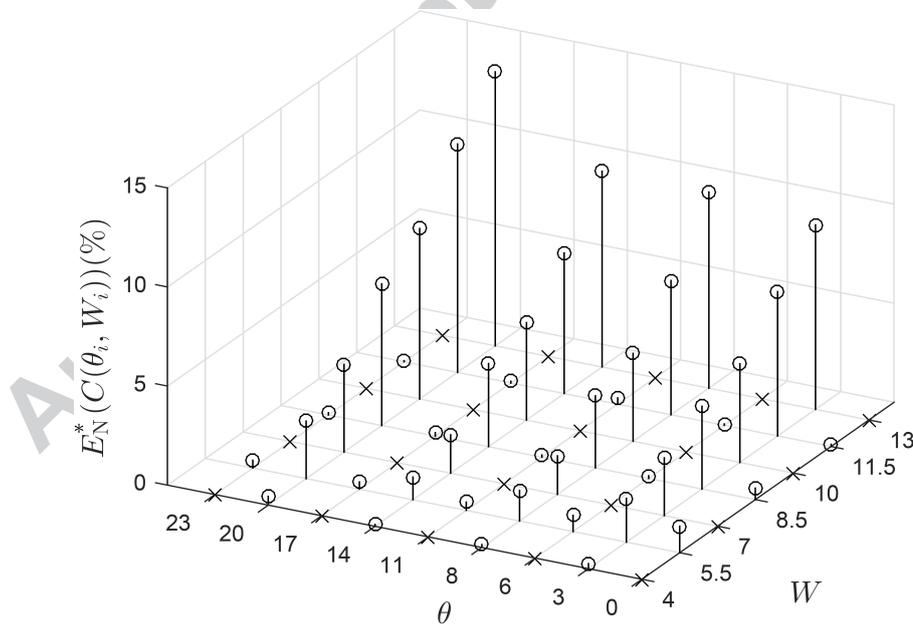
Fig. 5.7 Distribution of the errors obtained with the higher-order geodesic covariance family φ^{BG}. Crosses are training points, circles are test points. It is more difficult to recover the data when varying θ, particularly at larger W. Interpolation in the W direction, on the other hand, yields very small errors.

6 Conclusions

We have presented a differential geometric framework for constructing parametric low-rank covariance families, by connecting low-rank covariance matrices obtained at representative problem instances. In this sense, our framework creates parametric covariance families that can easily incorporate prior knowledge or empirical information, through the choice of the data matrices/anchors. We have proposed two main approaches: interpolation is performed either directly on the manifold, or on an affine section thereof. For each approach, we considered both multilinear, and patchwise cubic Bézier interpolation.

Given some data and the resulting sample covariance matrix (which can be strongly rank-deficient) we show how to perform minimum-distance covariance identification, that is, how to find the closest element of a given covariance family. We discuss methods and algorithms to solve this problem for each of the proposed covariance functions. In a case study involving wind velocity field approximation, we assess the ability of our covariance families to represent out-of-family covariance matrices. In practice, this technique can be used for data compression, i.e., storing the parameters representing a particular matrix of interest in the family, instead of the matrix itself. Moreover, our case study shows that a family constructed from covariance matrices obtained at very different wind headings contains matrices that match, within a reasonable accuracy for this problem, intermediate headings. Finally, one of main advantages of the proposed framework is it reduced computational cost, resulting from the fact that the proposed techniques only involve manipulations of $n \times p$ low-rank factors, instead of the $n \times n$ initial matrices. As illustrated in the case study, though the dimension of the covariance matrices is high ($n = 3024$), we have truncated their rank to $r = 20$ while keeping a reasonable accuracy in our results.

Other interesting questions not addressed in this paper would be the use of different multivariate interpolation techniques on manifolds, beyond those considered in this paper, or certain statistical questions, i.e., statistical properties of the minimum distance estimator as a function of the sample size q, the matrix dimension n, and the chosen rank r. We defer such investigations to future work.

Implementation code

The code to generate covariance functions and to perform identification and interpolation is available at https://github.com/EMassart/covariance_fitting.

A Computation of the tools required for the variable projection method

We detail here the computations for the main steps of the variable projection method proposed in Section 4.2.2. Remember that the corresponding surface, $\varphi^{\mathrm{LG}}(t_1, t_2) = Y_\varphi(t_1, t_2) Y_\varphi(t_1, t_2)^\top$, where Y_φ is obtained by composition of two geodesics, respectively along the t_1 and t_2 variables:

$$Y_\varphi(t_1, t_2) := (1 - t_2) Y_{1-2}(t_1) + t_2 Y_{3-4}(t_1) Q(t_1)^\top,$$
$$Y_{1-2}(t_1) := (1 - t_1) Y_1 + t_1 Y_2 Q_{1-2}^\top,$$
$$Y_{3-4}(t_1) := (1 - t_1) Y_3 + t_1 Y_4 Q_{3-4}^\top.$$

A.1 Computation of the partial derivative with respect to t_1

Computing the partial derivative of the function $(t_1, t_2) \mapsto \varphi^{\mathrm{LG}}(t_1, t_2)$ with respect to t_1 yields:

$$\frac{\partial \varphi^{\mathrm{LG}}}{\partial t_1}(t_1, t_2) = \frac{\partial Y_\varphi}{\partial t_1}(t_1, t_2) \, Y_\varphi(t_1, t_2)^\top + Y_\varphi(t_1, t_2) \left(\frac{\partial Y_\varphi}{\partial t_1}(t_1, t_2) \right)^\top,$$

with

$$\frac{\partial Y_\varphi}{\partial t_1}(t_1,t_2) = (1-t_2)\dot{\breve{Y}}_{1-2}(t_1) + t_2\dot{\breve{Y}}_{3-4}(t_1)Q(t_1)^\top + t_2 Y_{3-4}(t_1)\dot{Q}(t_1)^\top.$$

The values of $\dot{\breve{Y}}_{1-2}(t_1)$ and $\dot{\breve{Y}}_{3-4}(t_1)$ are independent of t_1:

$$\dot{\breve{Y}}_{1-2}(t_1) = -Y_1 + Y_2 Q_{1-2}^\top, \qquad \dot{\breve{Y}}_{3-4}(t_1) = -Y_3 + Y_4 Q_{3-4}^\top, \qquad \forall t_1.$$

The value of $\dot{Q}(t_1)$ can be obtained as follows. Recall from the geodesic definition that $Q(t_1)$ is the orthogonal factor of the polar decomposition of the matrix $M(t_1) = Y_{1-2}(t_1)^\top Y_{3-4}(t_1)$, which means that there exists a symmetric positive definite matrix $H(t_1)$ such that $M(t_1) = H(t_1)Q(t_1)$. Then,

$$\dot{M}(t_1) = \dot{H}(t_1)Q(t_1) + H(t_1)\dot{Q}(t_1), \tag{A.1}$$

where $\dot{H}(t_1)$ is a symmetric matrix, and $\dot{Q}(t_1)$ is of the form $\dot{Q}(t_1) = \Omega(t_1)Q(t_1)$, with $\Omega(t_1)$ a skew-symmetric matrix. Right-multiplying this expression by $Q(t_1)^\top$ yields:

$$\dot{M}(t_1)Q(t_1)^\top = \dot{H}(t_1) + H(t_1)\Omega(t_1), \tag{A.2}$$

while left-multiplying the transpose of equation (A.1) by $-Q(t_1)$ yields:

$$-Q(t_1)\dot{M}(t_1)^\top = -\dot{H}(t_1) + \Omega(t_1)H(t_1). \tag{A.3}$$

Now, summing equations (A.2) and (A.3) yields:

$$\dot{M}(t_1)Q(t_1)^\top - Q(t_1)\dot{M}(t_1)^\top = H(t_1)\Omega(t_1) + \Omega(t_1)H(t_1).$$

As a result, the term $\dot{Q}_1(t_1)$ can be obtained by solving a Sylvester equation. Moreover, since $H(t_1)$ is always positive definite (except in the set of zero measure corresponding to low-rank matrices $Y_{1-2}(t_1)^\top Y_{3-4}(t_1)$), the solution to the Sylvester equation is unique ($H(t_1)$ and $-H(t_1)$ have no common eigenvalues).

A.2 Computation of $t_2^*(t_1)$

The first-order optimality condition

$$\left.\frac{\partial f}{\partial t_2}\right|_{(t_1 t_2^*(t_1))} = 0$$

implies that the optimal value $t_2^*(t_1)$ corresponding to an arbitrary value t_1 is the solution to a cubic equation:

$$s_1(t_1)t_2^3(t_1) + s_2(t_1)t_2^2(t_1) + s_3(t_1)t_2(t_1) + s_4(t_1) = 0, \tag{A.4}$$

with

$$s_1(t_1) = 2\text{tr}\left(R^2\right) = 2\sum_i\sum_j R_{ij}^2,$$

$$s_2(t_1) = 3\text{tr}(RS) = 3\sum_i\sum_j R_{ij}S_{ij},$$

$$s_3(t_1) = 2\text{tr}(RT) + \text{tr}(S) = 2\sum_i\sum_j R_{ij}T_{ij} + S_{ij}^2,$$

$$s_4(t_1) = 2\text{tr}(ST) = 2\sum_i\sum_j S_{ij}T_{ij}.$$

The matrices R, S, and T arising in those expressions are defined as:

$$R = Y_{1-2}Y_{1-2}^\top + Y_{3-4}Y_{3-4}^\top - \left(Y_{1-2}QY_{3-4}^\top + Y_{3-4}Q^\top Y_{1-2}^\top\right),$$

$$S = \left(Y_{1-2}QY_{3-4}^\top + Y_{3-4}Q^\top Y_{1-2}^\top\right) - 2Y_{1-2}Y_{1-2}^\top,$$

$$T = Y_{1-2}Y_{1-2}^\top - \widehat{C},$$

with all these matrices depending on t_1. Observe, however, that for a fixed value of t_1, the function $t_2 \to f(t_1,t_2)$ might not be convex; hence, the condition (A.4) might have several (up to three) real solutions. In that case, we compute the value of the cost function at those solutions, and we choose the one corresponding to the smallest value of f.

A.3 Gradient descent for the univariate cost function

We are now looking for the derivative of the cost function $\tilde{f}(t_1) = f(t_1, t_2^*(t_1))$, with respect to the variable t_1, in order to be able to apply a steepest descent method to that problem. Using the notation $\tilde{f} = F \circ \tilde{\varphi}^{\mathrm{LG}}$, with $\tilde{\varphi}^{\mathrm{LG}}(t_1) = \varphi^{\mathrm{LG}}(t_1, t_2^*(t_1))$, we have:

$$\dot{\tilde{f}}(t_1) = DF[\dot{\tilde{\varphi}}^{\mathrm{LG}}(t_1)] = 2\mathrm{tr}\left(\dot{\tilde{\varphi}}^{\mathrm{LG}}(t_1)(\tilde{\varphi}^{\mathrm{LG}}(t_1) - \widehat{C})^{\top}\right).$$

The derivative $\dot{\tilde{\varphi}}^{\mathrm{LG}}(t_1)$ is given by:

$$\dot{\tilde{\varphi}}^{\mathrm{LG}}(t_1) = \dot{Y}_{\tilde{\varphi}}(t_1)Y_{\tilde{\varphi}}(t_1)^{\top} + Y_{\tilde{\varphi}}\dot{Y}_{\tilde{\varphi}}(t_1)^{\top}.$$

Using the chain rule,

$$\dot{Y}_{\tilde{\varphi}}(t_1) = \frac{\partial Y_{\varphi}}{\partial t_1}(t_1, t_2^*(t_1)) + \frac{\partial Y_{\varphi}}{\partial t_2}(t_1, t_2^*(t_1))\dot{t}_2^*(t_1).$$

By the definition of $t_2^*(t_1)$, the term $\frac{\partial Y_{\varphi}}{\partial t_2}(t_1, t_2^*(t_1))$ is equal to zero. As a result, $\dot{Y}_{\tilde{\varphi}}(t_1) = \frac{\partial Y_{\varphi}}{\partial t_1}(t_1, t_2^*(t_1))$, which has been computed earlier.

References

1. Absil, P.A., Gousenbourger, P.Y., Striewski, P., Wirth, B.: Differentiable piecewise-Bézier interpolation on Riemannian manifolds. In: ESANN2016, pp. 95–100. Springer (2016)
2. Absil, P.A., Gousenbourger, P.Y., Striewski, P., Wirth, B.: Differentiable piecewise-Bézier surfaces on Riemannian manifolds. SIAM Journal on Imaging Sciences **9**(4), 1788–1828 (2016)
3. Bergmann, R., Gousenbourger, P.Y.: A variational model for data fitting on manifolds by minimizing the acceleration of a Bézier curve. arXiv preprint arXiv:1807.10090 (2018)
4. Bonnabel, S., Sepulchre, R.: Riemannian metric and geometric mean for positive-semidefinite matrices of fixed-rank. SIAM Journal on Matrix Analysis and Applications **31**(3), 1055–1070 (2009)
5. Boumal, N., Absil, P.A.: A discrete regression method on manifolds and its application to data on SO(n). In: IFAC Proceedings Volumes (IFAC-PapersOnline), vol. 18, pp. 2284–2289 (2011). DOI 10.3182/20110828-6-IT-1002.00542
6. Cai, T., Liu, W.: Adaptive thresholding for sparse covariance matrix estimation. Journal of the American Statistical Association **106**(494), 672–684 (2011)
7. Cai, T., Liu, W., Luo, X.: A constrained L1 minimization approach to sparse precision matrix estimation. Journal of the American Statistical Association **106**(494), 594–607 (2011)
8. Cococcioni, M., Lazzerini, B., Lermusiaux, P.F.: Adaptive sampling using fleets of underwater gliders in the presence of fixed buoys using a constrained clustering algorithm. In: In Proc. of OCEANS'15, Genova, Italy, May 18-21 (2015)
9. Conti, C., Cotronei, M., Sauer, T.: Full rank positive matrix symbols: interpolation and orthogonality. BIT Numerical Mathematics **48**(1), 5–27 (2008)
10. Cover, T.M., Thomas, J.A.: Elements of information theory. John Wiley & Sons (2012)
11. Cressie, N.: The origins of kriging. Mathematical geology **22**(3), 239–252 (1990)
12. Cressie, N.: Statistics for spatial data, vol. 4-5. Wiley Online Library (1992)
13. Cressie, N., Hawkins, D.M.: Robust estimation of the variogram: I. Journal of the International Association for Mathematical Geology **12**(2), 115–125 (1980)
14. Cressie, N., Huang, H.C.: Classes of nonseparable, spatio-temporal stationary covariance functions. Journal of the American Statistical Association **94-448**, 1330–1339 (1999)
15. Csiszár, I., Shields, P.C.: Information theory and statistics: A tutorial. Foundations and Trends® in Communications and Information Theory **1**(4), 417–528 (2004)
16. Doekemeijer, B., Boersma, S., Pao, L.Y., van Wingerden, J.W.: Ensemble Kalman filtering for wind field estimation in wind farms. In: 2017 American Control Conference (ACC), pp. 19–24. IEEE (2017)
17. Driscoll, J.C., Kraay, A.C.: Consistent covariance matrix estimation with spatially dependent panel data. Review of economics and statistics **80**(4), 549–560 (1998)
18. Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. Biostatistics **9**(3), 432–441 (2008)
19. Furrer, R., Genton, M.G., Nychka, D.: Covariance tapering for interpolation of large spatial datasets. Journal of Computational and Graphical Statistics **15**(3), 502–523 (2006)
20. Gallager, R.G.: Information theory and reliable communication, vol. 588. Springer (1968)
21. Gousenbourger, P.Y., Massart, E., Absil, P.A.: Data fitting on manifolds with composite Bézier-like curves and blended cubic splines. Journal of Mathematical Imaging and Vision pp. 1–27 (2018)
22. Gousenbourger, P.Y., Massart, E., Musolas, A., Absil, P.A., Jacques, L., Hendrickx, J.M., Marzouk, Y.: Piecewise-Bézier C1 smoothing on manifolds with application to wind field estimation. In: ESANN2017, pp. 305–310. Springer (2017)
23. Grohs, P.: Quasi-interpolation in Riemannian manifolds. IMA Journal of Numerical Analysis **33**, 849–874 (2013)

24. Guerci, J.R.: Theory and application of covariance matrix tapers for robust adaptive beamforming. IEEE Transactions on Signal Processing **47**(4), 977–985 (1999)

25. Guhaniyogi, R., Banerjee, S.: Multivariate spatial meta kriging. Statistics & probability letters **144**, 3–8 (2019)

26. Hinkle, J., Fletcher, P.T., Joshi, S.: Intrinsic polynomials for regression on Riemannian manifolds. Journal of Mathematical Imaging and Vision **50**(1), 32–52 (2014). DOI 10.1007/s10851-013-0489-5

27. Jayasumana, S., Hartley, R., Salzmann, M., Li, H., Harandi, M.: Kernel methods on Riemannian manifolds with Gaussian RBF kernels. IEEE transactions on pattern analysis and machine intelligence **37**(12), 2464–2477 (2015)

28. Journée, M., Bach, F., Absil, P.A., Sepulchre, R.: Low-rank optimization on the cone of positive-semidefinite matrices. SIAM Journal on Optimization **20**(5), 2327–2351 (2010). DOI 10.1137/080731359

29. Kacem, A., Daoudi, M., Ben Amor, B., Berretti, S., Alvarez-Paiva, J.C.: A novel geometric framework on Gram matrix trajectories for human nehavior understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI) (2018). DOI 10.1109/TPAMI.2018.2872564

30. Karniadakis, G., Sherwin, S.: Spectral/hp element methods for computational fluid dynamics. Oxford University Press (2013)

31. Kim, K.R., Dryden, I.L., Le, H.: Smoothing splines on Riemannian manifolds, with applications to 3D shape space. arXiv preprint arXiv:1801.04978 pp. 1–23 (2018)

32. Langelaan, J.W., Alley, N., Neidhoefer, J.: Wind field estimation for small UAVs. Journal of Guidance, Control, and Dynamics **34**(4), 1016–1030 (2011)

33. Langelaan, J.W., Spletzer, J., Montella, C., Grenestedt, J.: Wind field estimation for autonomous dynamic soaring. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) (2012)

34. Larrabee, T., Chao, H., Rhudy, M., Gu, Y., Napolitano, M.R.: Wind field estimation in UAV formation flight. In: American Control Conference (ACC), 2014, pp. 5408–5413. IEEE (2014)

35. Lawrance, N.R., Sukkarieh, S.: Simultaneous exploration and exploitation of a wind field for a small gliding UAV. AIAA Guidance, Navigation and Control Conference, AIAA Paper **8032** (2010)

36. Lawrance, N.R., Sukkarieh, S.: Path planning for autonomous soaring flight in dynamic wind fields. In: Robotics and Automation (ICRA), 2011 IEEE International Conference on, pp. 2499–2505. IEEE (2011)

37. Ledoit, O., Wolf, M.: A well-conditioned estimator for large-dimensional covariance matrices. Journal of multivariate analysis **88**(2), 365–411 (2004)

38. Ledoit, O., Wolf, M.: Non-linear shrinkage estimation of large-dimensional covariance matrices. The Annals of Statistics **40**(2), 1024–1060 (2012)

39. Ledoit, O., Wolf, M.: Optimal estimation of a large-dimensional covariance matrix under Stein's loss. Bernoulli **24**(4B), 3791–3832 (2018)

40. Li, S.Z.: Markov random field modeling in computer vision. Springer Science & Business Media (2012)

41. Li, X.B., Burkowski, F.J.: Conformational transitions and principal geodesic analysis on the positive-semidefinite matrix manifold. In: Basu M., Pan Y., Wang J. (eds) Bioinformatics Research and Applications. ISBRA 2014. Lecture Notes in Computer Science, vol. 8492, pp. 334–345. Springer Cham (2014). DOI 10.1007/978-3-319-08171-7_30

42. Liu, J., Han, J., Zhang, Z.J., Li, J.: Target detection exploiting covariance matrix structures in MIMO radar. Signal Processing **154**, 174–181 (2019)

43. Lolla, T., Haley Jr, P., Lermusiaux, P.: Path planning in multi-scale ocean flows: Coordination and dynamic obstacles. Ocean Modelling **94**, 46–66 (2015)

44. Massart, E., Absil, P.A.: Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices. SIAM Journal on Matrix Analysis and Applications **41**(1), 171–198 (2020)

45. Massart, E., Gousenbourger, P.Y., Nguyen, T.S., Stykel, T., Absil, P.A.: Interpolation on the manifold of fixed-rank positive-semidefinite matrices for parametric model order reduction: preliminary results. In: ESANN 2019, pp. 281–286. Springer (2019)

46. Moakher, M., Batchelor, P.G.: Symmetric positive-definite matrices: From geometry to applications and visualization. In: Visualization and Processing of Tensor Fields, pp. 285–298. Springer (2006)

47. Modin, K., Bogfjellmo, G., Verdier, O.: Numerical algorithm for C2-splines on symmetric spaces. arXiv preprint arXiv:1703.09589 (2018)

48. Musolas, A., Smith, S.T., Marzouk, Y.: Geodesically parameterized covariance estimation. arXiv preprint arXiv:2001.01805 (2020)

49. Oliver, D.S., Chen, Y.: Recent progress on reservoir history matching: a review. Computational Geosciences **15**(1), 185–221 (2011)

50. Oliver, D.S., Cunha, L.B., Reynolds, A.C.: Markov chain Monte Carlo methods for conditioning a permeability field to pressure data. Mathematical Geology **29**(1), 61–91 (1997)

51. Palanthandalam-Madapusi, H.J., Girard, A., Bernstein, D.S.: Wind field reconstruction using flight data. In: 2008 American Control Conference, pp. 1863–1868. IEEE (2008)

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

53. Rasmussen, C.E.: Gaussian processes for machine learning. Citeseer (2006)

54. Ripley, B.D.: Spatial statistics, vol. 575. John Wiley & Sons (2005)

55. Rudovic, O., Pavlovic, V., Pantic, M.: Multi-output laplacian dynamic ordinal regression for facial expression recognition and intensity estimation. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2634–2641. IEEE (2012)

56. Rue, H., Held, L.: Gaussian Markov random fields: theory and applications. CRC press (2005)
57. Samir, C., Absil, P.A., Srivastava, A., Klassen, E.: A gradient-descent method for curve fitting on Riemannian manifolds. Foundations of Computational Mathematics **12**(1), 49–73 (2012). DOI 10.1007/s10208-011-9091-7
58. Sander, O.: Geodesic finite elements of higher order. IMA Journal of Numerical Analysis **36**, 238–266 (2016)
59. Schafer, J., Strimmer, K.: A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Applications in Genetics and Molecular Biology **4**(1), 1175–1189 (2005)
60. Srivastava, A., Klassen, E.P.: Functional and shape data analysis, vol. 1. Springer (2016)
61. Stein, M.L.: Interpolation of spatial data: some theory for kriging. Springer Science & Business Media (2012)
62. Stein, M.L., Chi, Z., Welty, L.J.: Approximating likelihoods for large spatial data sets. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **66**(2), 275–296 (2004)
63. Szczapa, B., Daoudi, M., Berretti, S., Bimbo, A.D., Pala, P., Massart, E.: Fitting, Comparison, and Alignment of Trajectories on Positive Semi-Definite Matrices with Application to Action Recognition. In: Human Behavior Understanding, satellite workshop of the International Conf. on Computer Vision 2019 (ICCV2019), arxiv:1908.00646 (2019)
64. Vandereycken, B., Absil, P.A., Vandewalle, S.: Embedded geometry of the set of symmetric positive-semidefinite matrices of fixed-rank. In: IEEE/SP 15th Workshop on Statistical Signal Processing, pp. 389–392 (2009). DOI 10.1109/SSP.2009.5278558
65. Vandereycken, B., Absil, P.A., Vandewalle, S.: A Riemannian geometry with complete geodesics for the set of positive-semidefinite matrices of fixed-rank. IMA Journal of Numerical Analysis p. drs006 (2012)
66. Wolfowitz, J.: The minimum distance method. The Annals of Mathematical Statistics **28**(1), 75–88 (1957)
67. Yang, S., Wei, N., Jeon, S., Bencatel, R., Girard, A.: Real-time optimal path planning and wind estimation using Gaussian process regression for precision airdrop. In: 2017 American Control Conference (ACC), pp. 2582–2587. IEEE (2017)