EXECUTIVE SUMMARY OF THE THESIS

Nonparametric estimation of spatial covariance functions for spatial prediction of multi-temporal DInSAR data

LAUREA MAGISTRALE IN MATHEMATICAL ENGINEERING - INGEGNERIA MATEMATICA

Author: ROBERTA TROILO

Advisor: PROF. ALESSANDRA MENAFOGLIO

Co-advisors: PROF. SIMONE VANTINI, TERESA BORTOLOTTI

Academic year: 2022-2023

1. Introduction and Preliminaries

In this thesis, we propose a nonparametric covariance reconstruction technique, rooted in the principles of functional data analysis (FDA), for the analysis of high-resolution images recording the temporal series of ground surface deformation registered by satellites. The images result from the processing of syntetic aperture radar (SAR) data via the Small Baseline Subsets methodology (SBAS, Berardino et al. [2002] [1]), a multi-temporal differential interferometric technique that provides ground deformation measurements over time with centimeter to millimiter accuracy. The presence of missing values in the SBAS-processed images is due to the scattering, absorption or reflection away from the sensor of radar signals, happening when specific spatial entities are radiated, such as water, vegetation, or rocks. This results in consistent missing values (*coherence* < 0.8) at the same locations across multiple temporal images. In the field of remote sensing, the reconstruction of these images at missing locations holds crucial significance in the endeavor to mitigate natural hazards. Our focus is on the area of Phlegrean Fields (Italy), reported in Figure 2.

In the context of FDA (Ramsay and Silvermann [2005] [6]), the issue of reconstructing missing data is essential. Each functional datum X_i, with $i = 1, ..., n$, represents one observation - among n realizations - of a random function $X(t) : D \to \mathbb{R}$, being D its domain of definition. When some of the realizations are only observed on a subset O_i of the domain, the available data are incomplete and most statistical methods designed for analysing functional data cannot be applied.

Specifically, the covariance operator estimator, that is typically employed in the reconstruction of partially observed functional data, may exhibit inconsistency, due to the unavailability of a sufficient number of curves being observed for every pair of domain locations, or even incompleteness, if no curve is observed for some couple of domain observations.

Seeking to address this issue, Descary and Panaretos [2018] [2] propose a non-parametric reconstruction of the covariance function, considering the domain $D = [0, 1]$ and assuming that each functional sample X_i is only observed on a generic subinterval of the domain $O_i \subset [0, 1]$, of fixed length δ, i.e. $\delta = |O_i| \, \forall i = 1, ..., n$,

with $0 < \delta < 1$. This regime, characterized as *banded* by the authors, stands apart from the *blanket* regime described in the work of Kraus [2015] [3]. In the blanket regime, O_i represents a union of subintervals rather than an interval itself, and the overall pattern of missing information is assumed to be such that there is sufficient information to get a consistent estimate of the covariance kernel over its entire domain. Conversely, Descary and Panaretos [2018], building upon the covariance kernel estimator of the work of Kraus [2015], develop a method for covariance reconstruction. Handling a different configuration for missing data, this work confronts the case in which each data sample is only registered on the same subset $O \subset D$ of the domain, differently from the banded regime of Descary and Panaretos [2018]. Indeed, in our problem setting $O_i = O, \forall i = 1, ..., n$, where O is not assumed to be connected, which implies the presence of completely unobserved rows and columns in the estimated covariance kernel. We define our regime as *fragmented*. Our aim is to find a method to reconstruct the covariance in the fragmented regime. With a complete covariance operator estimator and mean function estimator at hand, the approach outlined in Kraus [2015] introduces a functional completion technique to estimate the missing segments of each incomplete functional sample based on its observed portions.

1.1. Problem statement and notation

Let $X_1, \ldots, X_n$ be iid realizations of a random function X with values in the separable Hilbert space of square integrable functions on a bounded domain. Without loss of generality, this space is set as $L^2([0,1])$, with inner product $\langle f, g \rangle = \int_0^1 f(t)g(t)dt$, $f, g \in L^2([0,1])$ and $\|f\| = \langle f, g \rangle^{\frac{1}{2}}$. In the case of partially observed functional data, instead, the realizations are not observed in the whole domain, but in a subinterval O_i of it, such that $t \in O_i \subseteq [0,1]$, for all $X_i(t) \in \mathbb{R}$. As a consequence, $X_i(t)$ can be written as $X_i(t) = X_{iO_i}(t)\mathbf{1}_{O_i} + X_{iM_i}(t)\mathbf{1}_{M_i}$, where $M_i = [0,1]\backslash O_i$. The associate mean function is $\mu : [0,1] \to \mathbb{R}$, where $\mu(t) = \mathbb{E}[X_1(t)]$ a.e. for $t \in [0,1]$ and the covariance operator is $\mathcal{R} : L^2([0,1]) \to L^2([0,1])$ defined as:

$$\mathcal{R}f = \mathbb{E}[\langle f, X_1 - \mu \rangle (X_1 - \mu)] = \int_0^1 r(\cdot, t)f(t)dt$$

for any $f \in L^2([0,1])$, with $r(s,t)$ being the covariance kernel of the random function X_1, such that $r(s,t) = Cov(X_1(s), X_1(t))$ a.e. for $s, t \in [0,1]$, and for any $g, f \in L^2([0,1])$. For our purposes, it is important to estimate the mean function and the covariance operator. The estimation of the former can be readily obtained by considering the sample mean for each domain location. However, for locations where no observation is available, alternative techniques need to be employed. On the other hand, the estimate of the covariance operator is directly derived from the estimate of its kernel. Starting from the definitions of covariance kernel estimator of Kraus [2015] and Descary and Panaretos [2018], we define our covariance estimator $r_n(s,t)$ as

$$r_n(s,t) =$$
$$\frac{I(s,t)}{n}\left[\sum_{i=1}^n (X_i(s) - \mu_n(s))(X_i(t) - \mu_n(t))\right]$$

where $(s,t) \in [0,1]^2$, $I(s,t) = \sum_{i=1}^n U_i(s,t)$, where $U_i(s,t) = \mathbb{1}_{O_i}(s)\mathbb{1}_{O_i}(t)$ and $\mu_n(t)$ is the mean estimator evaluated at location t. As a result, $r_n(s,t) = 0$ if $s, t \in M = [0,1]\backslash O$, with $M_i = M$ for all $i = 1, ...n$, and these correspond to locations where the covariance has to be reconstructed.

We consider the case in which functional data are only measured on a finite grid of points. Even if the data in our case study are georeferenced, we stick to a one-dimensional representation of it. In Section 4, in fact, we motivate the choice of representing each two-dimensional functional datum as a one-dimensional object, while preserving the continuity across columns. Consequently, we represent each functional datum by its K evaluations at specific locations $(t_1, ..., t_K) \in T \subset O$, each corresponding to one successive point of the domain. As a consequence, the discretized covariance kernel can be synthesized by the $K \times K$ covariance matrix $R_n^K = \{r_n(t_j, t_l)\}_{j,l=1}^K$. Our purpose is to estimate R_n^K at its missing cells, such that the method of Kraus [2015] can be subsequently employed for functional completion.

1.2. Low-rank matrix completion

In the work of Descary and Panaretos [2018], in order to the estimate the covariance matrix in the banded regime, a matrix completion problem

is employed, which essentially aims to recover a low-rank matrix from a partially observed matrix. To ensure identifiability, the true covariance matrix R^K is assumed to admit a finite-rank Mercer decomposition of the true covariance matrix R^K, with real and analytic eigenfunctions. The rank minimization problem to be solved is

$$\min_{\theta \in \mathbb{R}^{K \times K}} \left\{ \frac{\|P^K \circ (R_n^K - \theta)\|_F^2}{K^2} + \tau \mathrm{rank}(\theta) \right\}$$

(1)

where $\tau > 0$ is a sufficiently small tuning parameter and P^K is a matrix with only 0 and 1 entries, having value 1 in the cells of the matrix corresponding to observed locations, 0 otherwise. By performing the element-by-element multiplication of P^K with $(R_n^K - \theta)$, one obtains the same $(R_n^K - \theta)$ matrix, filtered on the only observed part of the matrix, so assigning zero values to the missing cells. The practical problem resolution (Algorithm 1) consists in solving a series of rank-constrained minimization problems, with the addition of a hyperparameter $\tau > 0$ that prevents us from overfitting.

Algorithm 1 Best rank estimation algorithm

1: **for** $i = 1, ..., \lceil K\delta \rceil$ **do**
2:    solve the minimization problem

$$\min_{0 \leq \theta \in \mathbb{R}^{K \times K}} \left\{ \frac{\|P^K \circ (R_n^K - \theta)\|_F^2}{K^2} \right\}$$

subject to $\mathrm{rank}(\theta) \leq i$

3:    given the result $\hat{\theta}_i$ of the previous step,
     compute $f(i) = \frac{\|P^K \circ (R_n^K - \hat{\theta}_i)\|_F^2}{K^2}$
4: **end for**
5: choose the best rank i^* as the one minimizing $f(i) + \tau i$, for a fixed choice of $\tau > 0$.

As a consequence, by removing the rank constraint from the objective function, the problem achieves convergence in a much faster and simpler way. The choice of the value for the hyperparameter τ is finalized by plotting, for each possible value of τ, the solutions $f(i)$ of step 2 of the Algorithm 1 over the range of the rank values i. A non-increasing behaviour is expected in each plot, so the best τ is selected by observing an elbow in the plot and by demanding that $f(i_\tau)$ is lower than a certain threshold ϵ, where i_τ is the rank minimizing the objective function for that choice of τ.

2. Covariance estimation

Although the same assumptions of finite rank and analyticity can be made, the covariance kernel structure in the fragmented regime is not well-suited for the low-rank matrix completion procedure. For this reason, we propose an extension to their method to also account for the fragmented regime. To begin with, the method of Descary and Panaretos [2018] is not directly applicable to R_n^K in our fragmented regime. The algorithm depends on a specific continuity along the diagonal, which is evidently not preserved in our problem scenario. To address this, we suggest estimating the values of missing pixels along the diagonal of R_n^K to maintain continuity. Various techniques can be applied for this purpose, although we do not delve into them here for brevity. Once the diagonal is fully estimated, it is possible to apply Algorithm 1, which allows us to find the optimal rank for our desired resulting covariance matrix. Nevertheless, the application of the algorithm results in an irregular reconstruction of the covariance kernel. In fact, while the values of the observed cells are accurately estimated by well-preserving the information carried by the observed parts of the functional curves, the values of the missing cells are barely optimized in the minimization problem. As a consequence, we propose to conduct a new optimization problem to efficiently and entirely estimate the covariance kernel. In practice, we consider a fixed-rank matrix completion problem, with rank equal to the optimal rank, with an additional Laplacian regularization.

The inclusion of this latter additional term in the objective function to be minimized allows us to generate a much smoother reconstructed matrix. In fact, Laplacian regularization encourages smoothness in the solutions by penalizing abrupt changes or oscillations. This is particularly beneficial in problems involving data with spatial or relational dependencies, where a smooth behaviour is desired.

Additionally, we introduce weights in the Laplacian regularization term, by assigning varying weights to the Laplacian computed in the ob-

served or in the missing part of the covariance function, assuming that different parts of R_n^K might be weighted in different ways.

The result is the following regularization term:

$$\text{Tr}((P^m \circ (L^{\frac{1}{2}}\theta))^T (P^m \circ (L^{\frac{1}{2}}\theta))). \quad (2)$$

In (2), $\theta \in \mathbb{R}^{K \times K}$ is the discretized covariance kernel that is being reconstructed. $P^m \in \mathbb{R}^{K \times K}$ is the weight matrix, such that $P^m = (1 - m)P^K + mP^{-K}$, where $P^{-K} \in \{0,1\}^{K \times K}$ is the complementary mask. It assigns weight m to the missing cells and weight $1 - m$ to the observed cells. $L \in \mathbb{R}^{K \times K}$ is the Laplacian matrix (Maunu [2023] [4] and Pang [2017] [5]). It holds that $L = D - A$, where A is the adjacency matrix and D the degree matrix. $A \in \{0,1\}^{K \times K}$ is a symmetric matrix indicating whether or not two points in the domain - indexed from 1 to K - are connected, i.e. if j and l are connected, then $a_{jl} = 1$, otherwise $a_{jl} = 0$. $D \in \mathbb{R}^{K \times K}$ is a diagonal matrix indicating how many indices $l = 1, ..., K$, with $l \neq j$, are connected to each index j, i.e. such that $d_{jj} = \sum_{l=1}^{K} a_{jl}$.

The final optimization problem that we solve is

$$\min_{\theta \in \mathbb{R}^{K \times K}} \left\{ \|P^K \circ (R_n^K - \theta)\|_F^2 \right.$$
$$\left. + \alpha \text{Tr}((P^m \circ (L^{\frac{1}{2}}\theta))^T (P^m \circ (L^{\frac{1}{2}}\theta))) \right\} \quad (3)$$

with the constraint that $\text{rank}(\theta) = r$, with r optimal rank. The values of the hyperparameters $\alpha > 0$ and $m \in [0, 1]$ are fixed at this stage of the procedure. Hyperparameters α and m may be tuned according to some criteria, which we identify drawing on the analysis conducted in the simulation study and in the case study. In order to achieve a more rapid resolution, both problems (1) and (3) are solved by exploiting the positive-semidefiniteness of covariance matrices. In particular, θ is reparametrized as $\theta = \gamma \gamma^T$, with $\gamma \in \mathbb{R}^{i \times K}$. Moreover, the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm is used for the optimization in both cases.

The resolution of the minimization problem (3) results in finding a covariance matrix which effectively exploits the information carried by the observed cells, while guaranteeing continuity over the missing rows and columns of the matrix. Figure 1 reports a simulated covariance reconstruction for specific choice of parameters.
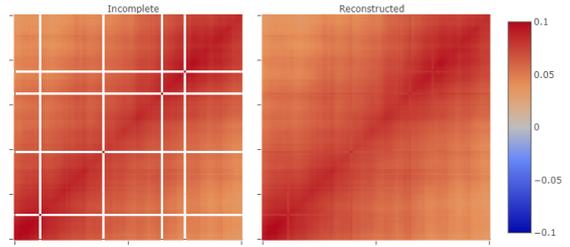

Figure 1: Reconstruction of the simulated covariance ($\alpha = 0.01$, $m = 0.13$).

3.   Simulation study

The simulation study is conducted with the aim of testing the performance of our method and, specifically, identifying the selection criteria for the hyperparameters α and m. Our simulated data consist of 100 functional samples of a Gaussian process having zero mean and Matérn covariance kernel. We generate realizations for each functional sample X_i over 101 locations, i.e. each functional datum is a curve over 101 successive equidistant points of the domain, and assign missing values at indices 12, 40, 66 and 76. Afterwards, we compute the covariance between each couple of locations (s, t). As a result, the covariance matrix is empty at rows and columns corresponding to the missing indices, since these are locations that are never observed in any of the 100 samples, including the diagonal values. Taking advantage of the stationarity of the Matérn covariance kernel, we set the missing values of the diagonal equal to the mean of the observed elements of the diagonal.

To test the performance of our method, we conduct a Monte Carlo simulation over 50 covariance functions, such that each covariance is computed from 100 samples of a centred Gaussian process with Matérn covariance function and displays the same pattern of missing cells. In order to compare the performance of different combinations of parameters α and m, we use the root mean squared error (RMSE) along all the cells of the matrix as a reconstruction index. Moreover, we separately compute RMSE over the observed cells and RMSE over the missing cells. Given the true covariance $R^K = \{r_{ij}^K\}_{i,j=1}^K$ and the reconstructed covariance $\hat{R}^K = \{\hat{r}_{ij}^K\}_{i,j=1}^K$, the root mean squared errors are formulated as follows:

$$\text{RMSE}(\hat{R}^K) = \sqrt{\frac{\sum_{i,j=1}^{K}(r_{ij}^K - \hat{r}_{ij}^K)^2}{K^2}}$$

$$\text{RMSE}_O(\hat{R}^K) = \sqrt{\frac{\sum_{i,j=1}^{K} P_{ij}^K (r_{ij}^K - \hat{r}_{ij}^K)^2}{N_{obs}}}$$

$$\text{RMSE}_M(\hat{R}^K) = \sqrt{\frac{\sum_{i,j=1}^{K} P_{ij}^{-K} (r_{ij}^K - \hat{r}_{ij}^K)^2}{N_{miss}}}$$

where N_{obs} is the number of non-zero elements of the mask matrix P^K and N_{miss} is the number of non-zero elements of the complementary mask P^{-K}, which is equal to the number of zero elements of P^K.

At this stage, the rank for each resulting covariance matrix is set to the best choice according to Algorithm 1 and covariances reconstructions are performed for several combinations of parameters α and m. We perform tuning of the parameters through grid search, selecting from values within the interval $[0, 1]$ for m and within the interval $[0, 100]$ for α, where α values are powers of 10. In Section 4 of the thesis, where the simulation study is conducted, the plots of mean RMSE over α values and over m values are reported, for the three types of error. We first consider the behaviour of the error with respect to the values of the logarithm of α, keeping m fixed. The curves of mean RMSE and RMSE_M attain a global minimum point for $\alpha = 0.01$. On the other hand, the curve of mean RMSE_O is constant for small values of α, and then suddenly starts increasing when $\alpha > 0.01$, i.e. the actual optimal α. Indeed, it is expected that, if α is large, the second term of the optimization problem described in (3) predominates over the first term. This suggests that, when the model attempts to estimate the covariance, the significance of the information provided by the observed cells is reduced, resulting in an increase in the error associated with the observed part. As a consequence, looking at the curve for RMSE_O, which is the only information available in a real data scenario, we can select the best value for α in correspondence of the elbow before the abrupt increase of the RMSE_O curve. This is not the case when considering the behaviour of the error with respect to the values of m. While, similarly to the previous case, the mean RMSE and RMSE_M exhibit a global minimum, the curve for RMSE_O is definitely decreasing, justified by

the fact that the observed part of the matrix is better estimated if only optimized over the first term of the objective function of the minimization problem (3). For this reason, another approach should be used to select m in real data analyses.

4. Case study

Finally, we focus on the application to the SBAS-processed images of the area of Phlegraen Fields, Italy. We test on real data our new covariance reconstruction method, drawing on the conclusions of the analysis carried out in simulation study. The area of interest for us has the size of 101×101 and comprises n=391 time observations. With the objective of reconstructing the incomplete temporal series of ground displacement, our initial step involves examining the independence among time instants. However, this condition is not met, as the temporal images clearly exhibit autocorrelation. Consequently, we preprocess our data to eliminate autocorrelation and retrieve independence within our problem framework.
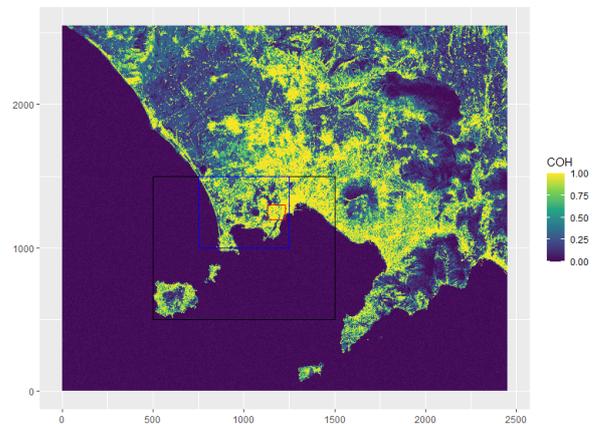

Figure 2: Coherence value per pixel in Phlegraean Fields, Italy

We begin our analysis by considering one single column 101×1 of the area of interest and estimating its covariance. We set the rank according to Algorithm 1 and choose α based on the criterion established in the simulation study, by looking at the curve of RMSE_O for some fixed values of m. At this stage, a criterion for the selection of m is needed. In order to find it, we conduct a cross-validation study to investigate various spatial configurations of missing rows and columns within the matrix. Specifically, we seek to exam-

ine whether the distance between missing rows and columns in the matrix impacts its reconstruction and, more importantly, influences the selection of a global minimum for the parameter m of the Laplacian regularization term. To perform the cross-validation in this setting, the idea is to subsequently consider the rows and columns near to the already missing rows and columns of the covariance kernel, and fictitiously treat them as missing. For each of these configurations, we consider the reconstruction of the matrix - at fixed (best) rank equal - for all the possible values of m and for optimal α and we compute the error made in reconstructing them. Examining the pattern of the reconstruction errors for the several configurations, the values for an optimal m are 0.05, 0.1 or 0.15. As a result, we can observe that the cross-validation study leads to achieving a global minimum for the reconstruction error across the range of values for m. As a consequence, we found a criterion for parameter selection for m.

To conclude, we consider the reconstruction of a two-dimensional surface. To accomplish this, we consider the transformation of each two-dimensional datum into a one-dimensional form, by keeping continuity across columns. This choice is strongly justified by the significant computational load associated with handling two-dimensional data in this type of problem, which would require a four-dimensional covariance. The relation between entire close rows and columns in the covariance is an information that would be lost by considering the reconstruction of one column at a time. Indeed, in that case, only the diagonal blocks would be available for data reconstruction. Nevertheless, not all the dependance is captured. In fact, when stacking columns vertically, although continuity is maintained, we lose information regarding the dependence between adjacent row elements. In this context, a potential expansion of our method would entail redefining the Laplacian term to take into consideration the four-dimensional proximity of elements within the covariance.

5. Conclusions

The significance of this thesis lies in the ability to effectively reconstruct the covariance kernel for partially observed functional data, present-

ing good performance in the associated functional completion. To attain this, we consider the low-rank matrix completion technique proposed by Descary and Panaretos [2018], which minimizes the rank of the covariance matrix and preserves the values of the observed cells, and add a Laplacian regularization term, that promotes smoothness across neighbouring cells. The performance of our method is showed in a simulated scenario, while its application for functional completion is shown in the case study section of the thesis. Moreover, the simulation study and the case study provide criteria for determining the most suitable regularization parameter α and weight parameter m, demonstrating their efficacy. Finally, by applying our method to reconstruct the temporal series of ground displacement, we successfully achieve functional completion.

References

[1] P. Berardino, G. Fornaro, R. Lanari, and E. Sansosti. A new algorithm for surface deformation monitoring based on small baseline differential sar interferograms. *IEEE Transactions on Geoscience and Remote Sensing*, 40(11):2375–2383, 2002.

[2] Marie-Eve Descary and Victor M Panaretos. Recovering covariance from functional fragments. *Biometrika*, 105(4):883–896, 2018.

[3] Daniel Kraus. Components and completion of partially observed functional data. *Journal of the Royal Statistical Society*, 77(4):777–801, 2015.

[4] T. Maunu. First-order algorithms for optimization over graph laplacians. In *2023 International Conference on Sampling Theory and Applications (SampTA)*, pages 1–11, New Haven, CT, USA, 2023. IEEE.

[5] J. Pang and G. Cheung. Graph laplacian regularization for image denoising: Analysis in the continuous domain. *IEEE Transactions on Image Processing*, 26(4):1770–1785, April 2017.

[6] James O Ramsay and Bernard W Silverman. *Functional Data Analysis*. Springer, 2005.