Statistical Analysis of Manifold-Valued Random Fields

Student: Savvas Melidonis Supervisor: Dr. Andrew Duncan

Department of Mathematics
Imperial College London

Project Presentation, September 2020

Table of Contents

Problem Overview

- A n-manifold is a topological space with the property that each point has a neighborhood that is homeomorphic to the Euclidean space of dimension n. Jost and Jost [2008].
- We study manifold-valued random fields with non-euclidean geometry attached to them.
- We concentrate on matrix-valued random fields.

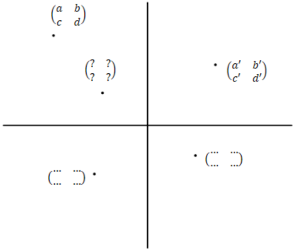

Figure: Problem Representation.

Examples of Matrix Manifolds

Example

The topological space of positive definite matrices is a manifold.

Example

The topological space of d- dimensional orthonormal bases of $\mathbb{R}^n$, $d \leq n$, is a manifold known as Grassmann manifold. It is strongly connected to dimension reduction applications (PCA and related techniques).

Previous Work

- The statistical analysis of non-linear data is not an uncommon task.
- In the work of Pigoli et al. [2016], the authors aim to model a matrix field of covariances including spatial dependence in the analysis.
- An additive linear model is proposed for ordinary kriging predictions on the manifold.

Motivation

- For the application in mind, generating a single point is very expensive, so the number of spatial points is small.
- The uncertainty of predictions is important.
- The above context motivates a Bayesian approach.

In this Project

- Our aim is to develop Bayesian spatial models for interpolating (kriging) matrix valued data.
- Initially, we focus on the manifold of positive definite matrices. Two models for interpolating positive definite matrix fields are proposed.
- Then, a model for interpolating random fields of orthonormal bases is proposed.
- We develop a Markov Chain Monte Carlo (MCMC) sampling method for posterior inference.

Table of Contents

Setting the Model

- The first model assumes a random field of non-negative definite matrices $W(s)$, where $s \in \mathbb{R}^k$.
- Suppose a dataset of the form $\mathcal{D} = \{(s_1, D_1), (s_2, D_2), \ldots, (s_N, D_N)\}$, where $s_1, \ldots, s_N \in \mathbb{R}^k$ are the spatial points and $D_1, \ldots, D_N$ are positive semi-definite matrices.

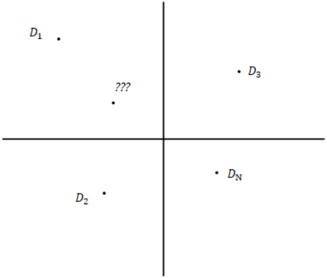

Figure: Random field representation.

Prior over the Matrix Field

- In the paper of Wilson and Ghahramani [2010], the authors introduce the definition of a Generalised Wishart Process (GWP).
- Assuming a Generalised Wishart Process over the matrix field means that every matrix W_j is expressed as

$$W(s_j) = \sum_{i=1}^{p} L u^{(i)}(s_j) u^{(i)}(s_j)^T L^T \ \ \forall j = 1, \ldots, N \ ,$$

  where L is a scale matrix such that $LL^T = V$, $p \in \mathbb{N}$ are the degrees of freedom and $u^{(i)} \in \mathbb{R}^D$ are vectors generated from a multivariate Gaussian Process.
- Typically, each matrix W_j follows marginally a Wishart distribution.

Likelihood over the Data

We propose a likelihood of the form :

$$p(D_j|u_1(s_j),\dots,u_p(s_j),L,p,\epsilon) \propto \exp\left(-\frac{||D_j - \sum_{j=1}^{p} Lu^{(i)}(s_j)u^{(i)}(s_j)^T L^T||_F^2}{2\epsilon}\right) .$$

where the term ϵ can be seen as a linear normal error.
Assuming that the matrix-valued variables D_j are independent, we have that

$$p(\mathcal{D}|\boldsymbol{u},L,p,\epsilon) \propto \prod_{j=1}^{N} p(D_j|u^{(1)}(s_j),\dots,u^{(p)}(s_j),L,p,\epsilon) ,$$

where $\boldsymbol{u} \equiv \mathrm{vec}([u^{(1)}(s_1),\dots,u^{(p)}(s_1),\dots,u^{(1)}(s_N),\dots,u^{(p)}(s_N)])$.

Posterior Distribution

Having the prior and the likelihood defined, we can derive the posterior distribution over the vectors

$$u^{(1)}(s_1), \ldots, u^{(p)}(s_1), \ldots, u^{(1)}(s_N), \ldots, u^{(p)}(s_N) \ .$$

Specifically, we have that

$$p(\boldsymbol{u}|\mathcal{D}, L, p, \ell) \propto p(\mathcal{D}|\boldsymbol{u}, L, p) \cdot p(\boldsymbol{u}|\ell) \ .$$

where $\boldsymbol{u} \equiv \text{vec}([u^{(1)}(s_1), \ldots, u^{(p)}(s_1), \ldots, u^{(1)}(s_N), \ldots, u^{(p)}(s_N)])$ is generated from a multivariate Gaussian distribution.

Bayesian Model

- Priors are considered over some unknown parameters of the model.
- These include the measurement error ϵ, the scale matrix $\boldsymbol{L}$ and the hyperparameter ℓ of the kernel function of the multivariate Gaussian Process.
- We make use of the Gibbs Sampling to sample in cycles from the posteriors of $\boldsymbol{u}$, ϵ, $\boldsymbol{L}$ and ℓ.
- We are based on Gaussian random walks to propose new states. We make use of Metropolis-Hastings algorithm to accept or reject the proposals.

Simulations

- We downloaded data from the Ogimet website. It is a Weather Information Service that provides meteorological (Synop) data.
- The measurements are vectors which include the daily average temperature and the daily precipitation totals from 10 different stations.
- For each station, the sample covariance matrix of these vectors is calculated.

Figure: Stations and observations.

Results

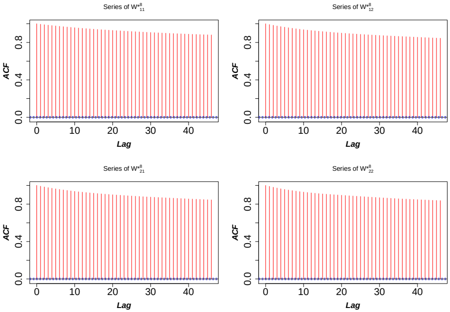

Figure: Autocorrelation plots of the Markov chains of all the components of the matrices at point s_8.

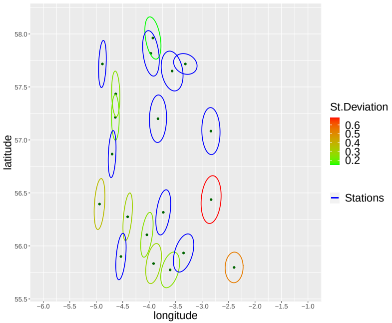

Figure: Representation of the predicted matrices in February.

Results Second Model

Table of Contents

Motivation for a Geometric Approach

- The previous model assumed a non-negative definite matrix at each point. The interpolated matrix field may attain some degenerate matrices with zero eigenvalues.
- The assumption of a normal linear error in a model that interpolates on a non-linear space.

These weaknesses **motivated** the construction of a model that interpolates on the **tangent space** of the manifold of positive definite matrices which is **linear**.

Manifold's Tangent Space

- The data matrices are transferred from the space $Sym^{+}(D)$ of positive definite matrices of dimension D to the tangent space.
- The tangent space $T_{\Sigma}Sym^{+}(D)$ in the point $\Sigma \in Sym^{+}(D)$ can be identified with the linear space $Sym(D)$ of symmetric matrices. [Pigoli et al., 2016]

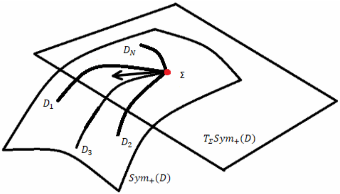

Figure: Transferring to tangent space.

A Geometric Approach

- The space $Sym^+(D)$ is a Riemannian manifold.
- An inner product $\langle \cdot\, , \cdot \rangle$ can be equipped on the tangent space $T_\Sigma Sym^+(D)$.
- For any pair $(U, \Sigma) \in Sym(D) \times Sym^+(D)$, there is a unique geodesic $g_\Sigma(t; U)$ with $g_\Sigma(0; U) = \Sigma$ and $\dot{g}_\Sigma(0; U) = U$. [Yuan, Zhu, Lin, and Marron, 2012]
- Based on the geodesics, there is a one-to-one correspondence between $Sym(D)$ and $Sym^+(D)$:

$$\exp_\Sigma : T_\Sigma Sym^+(D) \to Sym^+(D) \;\; , \;\; \exp_\Sigma(U) \equiv g_\Sigma(1; U)$$

$$\log_\Sigma : Sym^+(D) \to T_\Sigma Sym^+(D) \;\; , \;\; \log_\Sigma(\cdot) = \exp_\Sigma^{-1}(\cdot)$$

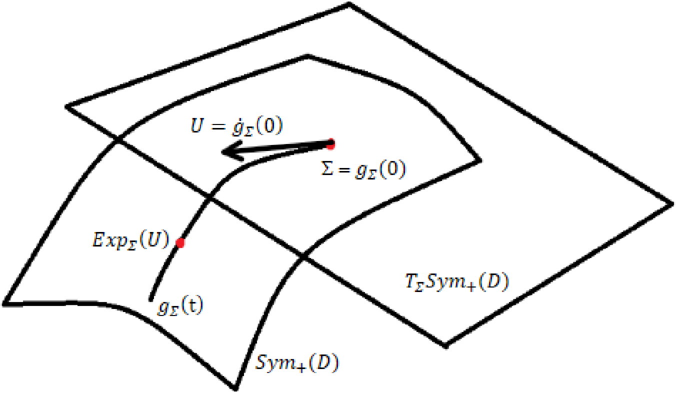

Figure: A geometrical representation.

The Updated Model

We update the likelihood as :

$$p(D_j|W(s_j)) \propto \exp\left(-\frac{||\log_{\Sigma}(D_j) - W(s_j)||_F^2}{2\epsilon}\right) .$$

- Now, the the matrices W_j should be modelled as generic symmetric matrices.
- We model the random field of W matrices as the difference of two Wishart Processes :

$$W(s_j) = W_u(s_j) - W_v(s_j) = \sum_{i=1}^{p} Lu^{(i)}(s_j)u^{(i)}(s_j)^T L^T - \sum_{i=1}^{p} Lv^{(i)}(s_j)v^{(i)}(s_j)^T L^T .$$

Results

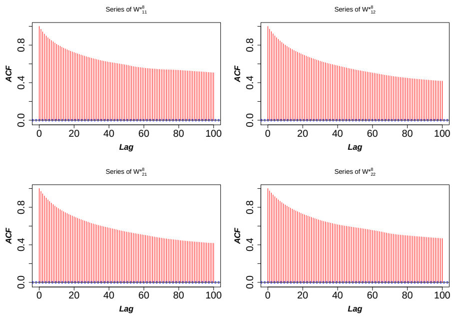

Figure: Autocorrelation plots of the Markov chains of all the components of the matrices at point s_8.

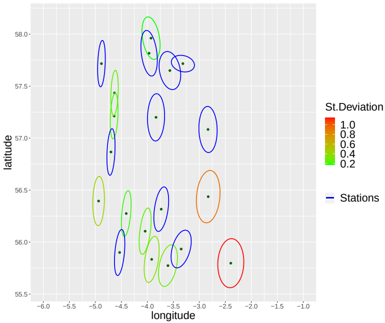

Figure: Representation of the predicted matrices in February.

▸ Results First Model

Table of Contents

Grassmann manifold

The set of all orthonormal bases of rank d forms the Grassmann manifold :

$$Gr(D,d) := \{[W] \mid W \in \mathbb{R}^{D\times d}\}.$$

The tangent space of $Gr(D,d)$ at a point U is defined as

$$T_{[U]}Gr(D,d) = \{\Delta \in \mathbb{R}^{D\times d} \mid U^T\Delta = 0\} \subset \mathbb{R}^{D\times d}.$$

We equip the tangent space $T_{[U]}Gr(D,d)$ with a positive inner product.
The inner product introduces a Riemannian exponential mapping

$$Exp^{Gr}_{[U]} : T_{[U]}Gr(D,d) \to Gr(D,d)$$

and a Riemannian logarithm mapping

$$Log^{Gr}_{[U]} : Gr(D,d) \to T_{[U]}Gr(D,d) \ .$$

Motivation

- At each $s_1, \ldots, s_N$, we appoint high-dimensional real symmetric matrices $C_1, \ldots, C_N \in \mathbb{R}^{D \times D}$.
- We acquire the first $d << D$ *important* eigenvectors $w_j^1, \ldots, w_j^d$ of matrix $C(s_j)$. They represent the most of the information. We form the respective matrices $W_j = [w_j^1, \ldots, w_j^d] \in \mathbb{R}^{D \times d}$.
- The column space $\mathcal{W}_j \subset \mathbb{R}^D$ of the respective matrix W_j is commonly a subspace such that $\dim(\mathcal{W}_i) = d \leq D$.

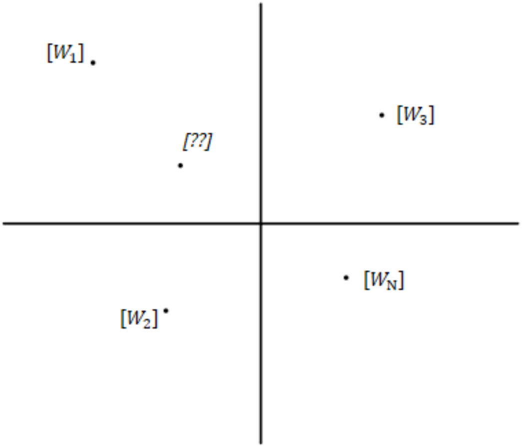

Figure: Application representation.

Prior Distribution over Low Rank Matrices

Suppose a random field of low rank orthonormal bases

$$[V(s)] = [v^1(s), \dots, v^d(s)] \in \mathbb{R}^{D\times d} \ ,$$

where $s \in \mathbb{R}^2$. We assume a multivariate Gaussian Process over their column vectors

$$\{v^i(s_j)\}_{i=1}^{d} \ , \forall j = 1, \dots, N \ .$$

Equivalently, the vectorization

$$\mathrm{vec}([v^1(s_1), \dots, v^d(s_1), \dots, v^1(s_N), \dots, v^d(s_N)]) = \boldsymbol{v}$$

has a joint multivariate Gaussian distribution.

Likelihood Proposal

A tangent vector $\Delta \in T_{[U]}Gr(D,d)$ can be represented [Zimmermann, 2019] as

$$\Delta = (I - UU^T)V\,, \quad V \in \mathbb{R}^{D\times d} \text{ arbitrary}\,, \quad I \in \mathbb{R}^{D\times D} \text{ the identity matrix.}$$

Similarly with the previous models, the likelihood is defined as

$$p(W_j|V_j) \propto \exp\left(-\frac{||(I - UU^T)V_j - Log_{[U]}(W_j)||_F^2}{2\epsilon}\right) \quad \forall j = 1,\dots,N\ .$$

Finally, the likelihood takes the form

$$p(\mathcal{D}|\boldsymbol{v},\epsilon) \propto \prod_{j=1}^{N} p(W_j|V_j,\epsilon)\ .$$

Sampling from the posteriors

- The unknown parameters of the model are the hyperparameter ℓ and the measurement error ϵ. We define lognormal priors over them.
- Gibbs Sampling is used to sample in cycles from the posteriors of $\boldsymbol{v}$, $\boldsymbol{\ell}$, $\boldsymbol{\epsilon}$. We use random walks combined with Metropolis-Hastings algorithm to propose new states.
- After we run the MCMC sampler, we acquire a random field on the tangent space. The exponential mapping can transfer the random field from the tangent space to Grassmann manifold.

Simulations

- A mesh of 114 spatial points $s_1, \ldots, s_N$ is assumed. Additionally, we add randomly 86 spatial points $s_1^*, \ldots, s_{N'}^*$ where predictions will be made.

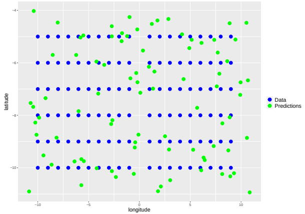

Figure: Mesh of spatial points.

- For these simulations, we have that $D = 10$ and $d = 1$. At each spatial point, we appoint an orthonormal matrix $W_j \in \mathbb{R}^{10\times 1}$. So we get a dataset of the form $\mathcal{D} = \{(s_1, [W_1]), (s_2, [W_2]), \ldots, (s_N, [W_N])\}$.

Results from Simulations

To represent the random field of orthonormal bases we need to implement:

- Generalised Procrustes Analysis (GPA)
- Principal Component Analysis (PCA)

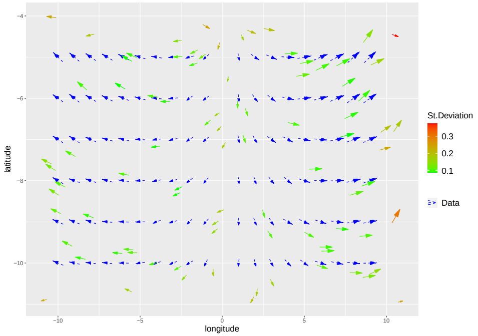

Figure: Representation of obtained random field.

Conclusion

- The proposed Bayesian models constitute a good first step for interpolating random fields of matrix-valued data.
- Slow convergence.
- Methods of variance reduction should be investigated.

References I

J. Jost and J. Jost. *Riemannian geometry and geometric analysis*, volume 42005. Springer, 2008.

D. Pigoli, A. Menafoglio, and P. Secchi. Kriging prediction for manifold-valued random fields. *Journal of Multivariate Analysis*, 145: 117–131, 2016.

A. G. Wilson and Z. Ghahramani. Generalised wishart processes. *arXiv preprint arXiv:1101.0240*, 2010.

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

R. Zimmermann. Manifold interpolation and model reduction. *arXiv preprint arXiv:1902.06502*, 2019.