POLITECNICO
MILANO 1863

Solving Large Mixed-Effects Models in Spatial Regression with PDE Regularization

M.Sc. Thesis of
Martino Ischia

Supervisor:
Prof. Laura Sangalli

Co-supervisor:
Dr. Eleonora Arnone

Programme:
Mathematical Engineering

Department of Mathematics
Politecnico di Milano

Contents

List of Figures

Abstract

aa

Sommario

Traduzione dell'abstract in italiano.

1. Synopsis

Here goes the synopsis of your thesis. It consists of an introductory part followed by a description on the thesis content.

This thesis is organized as follows.

2. Introduction

2.1. Spatial regression with partial differential equation regularization

This work naturally stems from the studies in the research field called *spatial regression with partial differential equation regularization*, abbreviated in the following as SR-PDE. SR-PDE constitutes a family of models which has been, and still is, under study beginning from 2013 in Sangalli, Ramsay, and Ramsay [SRR13].

SR-PDE effectively combines several branches of mathematics and statistics modeling. For a thorough analysis of its derivation and the enhancements it brings to the existing literature see [San21], I will limit myself to a brief introduction.
SR-PDE models feature the minimization of a functional composed of two terms: a classical least-square term, with the goal of estimating a vector of unknown regression coefficients (in case of presence of covariates), and a regularization term, estimating an unknown deterministic spatial field.
The spatial field contributes to the functional by integration over the spatial domain of the square of a differential operator, most commonly the Laplacian operator. By properties of the Laplace operator, this choice induces an isotropic and stationary smoothing effect on the unknown spatial field, meaning equal in every direction and independent on the position. Other differential operators considered in this work are generic second order differential operators, that come from the physics of the problem or from some preexisting knowledge about the phenomenon under study.

The inclusion of partial differential equations in the statistical model brings many advantages, like the aforementioned ability to include problem-specific knowledge, dealing with boundary conditions, etc., but makes computations more cumbersome.

2.2. Mixed-effects models

In a situation where observed data possess a natural grouping structure, mixed-effects models are often utilized. They combine fixed effects, meaning that a set of covariates have regression coefficients shared among all groups, and random effects, meaning the remaining set of covariates have regression coefficients varying along each group.

Given our interest in spatial data analysis, a typical situation in which mixed-models are used is in clinical data, where data of several patients are measured in the same area of the body. For example in chapter 6 Applications, we will see a mixed-effects model applied to human brain data, collected on a set of patients who undertook functional Magnetic Resonance Imaging (fMRI).

For this reason in the following we will use the word *patients* to indicate the different groups of our generic mixed-effects model.

We are therefore ready to present, similarly as in [Kim20], the SR-PDE mixed-effects model.

2.3. Generic SR-PDE mixed-effects model

Consider m patients. To the i-th patient, with i varying from 1 to m, corresponds a unique spatial domain Ω_i, on which we observe n_i data at different positions $\mathbf{p}_{ij}$. Index j is therefore varying from 1 to n_i, for given patient i.
The variable of interest z, observed at $\mathbf{p}_{ij}$, is modeled as

$$z_{ij} = \boldsymbol{w}_{ij}^T \boldsymbol{\beta} + \boldsymbol{v}_{ij}^T \mathbf{b}_i + f_i(\mathbf{p}_{ij}) + \epsilon_{ij}. \tag{2.1}$$

The notation used represents the following quantities:

- $\boldsymbol{w}_{ij}$ is the vector of fixed effects covariates for the observation z_{ij};
- $\boldsymbol{\beta}$ is the vector of regression coefficients for fixed effects;
- $\boldsymbol{v}_{ij}$ is the vector of random effects covariates for the observation z_{ij};
- $\mathbf{b}_i$ is the vector of regression coefficients of the random effects for patient i;
- f_i is the unknown deterministic field defined on domain Ω_i;
- ϵ_{ij}, for every i and j, are the random noise or errors, considered as the realizations of independent identically distributed (i.i.d.) random variables, with mean 0 and variance σ^2.

We will assume that the random effect $\mathbf{b}_i$ has average 0

$$\sum_{i=1}^{m} \mathbf{b}_i = 0. \tag{2.2}$$

In fact, if the average was different from 0, we would simply split the contribution of those covariates into two, a fixed effect one and a random effect covariate with 0 average across groups (more on this in section 2.4, Reparametrizing the model).
Let also N be equal to $\sum_{i=1}^{m} n_i$. For every i, j equation 2.1 constitutes a system of N equations, that is expressed concisely in algebraic form in the following two ways:

$$\begin{cases} \boldsymbol{z}_i = W_i \boldsymbol{\beta} + \mathbf{V}_i \mathbf{b}_i + \mathbf{f}_i + \boldsymbol{\epsilon}_i & i = 1 \dots m \\ \sum_{i=1}^{m} \mathbf{b}_i = 0 \end{cases} \tag{2.3}$$

and

$$\boldsymbol{z} = W\boldsymbol{\beta} + \mathbf{V}\mathbf{b} + \mathbf{f}_N + \boldsymbol{\epsilon}. \tag{2.4}$$

Describing each corresponding term in the two equations:

- $z_i \in \mathbb{R}^{n_i}$ is the vector of observed data for patient i and z is the vector belonging to $\mathbb{R}^N$ obtained by concatenating the m vectors z_i;
- W_i is the matrix whose element $w_{j,k}$ is the k-th element of previously defined vector $\boldsymbol{w}_{ij}$, $(\boldsymbol{w}_{ij})_k$, whereas W is a concatenation as well

$$W = \begin{bmatrix} W_1 \\ \vdots \\ W_m \end{bmatrix}; \tag{2.5}$$

- V_i is the matrix whose element $v_{j,k}$ is the k-th element of previously defined vector $\boldsymbol{v}_{ij}$, $(\boldsymbol{v}_{ij})_k$;
- when moving to the N system of equations, we include constraint 2.2 by defining $\mathbf{b}$ as the concatenation of $m-1$ vectors $\mathbf{b}_i$, omitting the last one

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_{m-1} \end{bmatrix} \tag{2.6}$$

  and defining V in the following manner:

$$V = \begin{bmatrix} V_1 & 0 & 0 & 0 \\ 0 & V_2 & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & 0 & V_{m-1} \\ -V_m & -V_m & -V_m & -V_m \end{bmatrix}; \tag{2.7}$$

- $\mathbf{f}_i \in \mathbb{R}^{n_i}$ is the vector of observed data for patient i and $\mathbf{f}_N$ is the vector belonging to $\mathbb{R}^N$ obtained by concatenating the m vectors $\mathbf{f}_i$;
- $\boldsymbol{\epsilon}_i \in \mathbb{R}^{n_i}$ is the vector of errors for patient i and $\boldsymbol{\epsilon}_N$ is the vector belonging to $\mathbb{R}^N$ obtained by concatenating the m vectors $\boldsymbol{\epsilon}_i$.

2.4. Reparametrizing the model

The N system of equations 2.4 is the model written in an unconstrained form, where the constraint 2.2 has been "injected" into the design matrix. We also call it *official parametrization* of the mixed-effects model.

As described in [Kim20] we prefer to adopt a slightly different approach, mainly for computational efficiency reasons. In the following we will assume that all covariates contribute to the fixed effect part of the model, whereas a subset of them will contribute to the random effect part. We will refer to this version of the model as the *implementation* one.

Implementation version is expressed by the following system:

$$z = W'\boldsymbol{\beta}' + V'\mathbf{b}' + \mathbf{f}_N + \boldsymbol{\epsilon}, \tag{2.8}$$

where we have used the following new quantities:

- W', similarly as before, is concatenation of W'_i matrices, where W'_i is composed of the covariates related only to fixed effects, meaning with no patient-specific contribution.

$$W' = \begin{bmatrix} W'_1 \\ W'_2 \\ \vdots \\ W'_m \end{bmatrix}; \tag{2.9}$$

- $\boldsymbol{\beta}'$ is the coefficient relative to covariates for fixed effects, as just described;

- V' is composed of V'_i matrices, which are the matrices of covariates having also a random effect. Unlike previous V, V' does not include the part deriving from constraint 2.2:

$$V' = \begin{bmatrix} V'_1 & 0 & 0 & 0 \\ 0 & V'_2 & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & 0 & V'_m \end{bmatrix}; \tag{2.10}$$

- $\mathbf{b}'$ is the regression coefficient relative to the covariates with patient-specific effect,

$$\mathbf{b}' = \begin{bmatrix} \mathbf{b}'_1 \\ \vdots \\ \mathbf{b}'_m \end{bmatrix}; \tag{2.11}$$

Let us also use the following notation, which allows us to properly refer to sizes of vectors and matrices defined before.
We indicate with q the number of covariates, and with p the number of covariates with patient-specific interest. Therefore $q - p$ is the number of covariates considered for their fixed effect contribution only.

Vector $\mathbf{b}'$ belongs to $\mathbb{R}^{mp}$, $\boldsymbol{\beta}'$ to $\mathbb{R}^{q-p}$ and we can relate it to previously defined vectors $\mathbf{b}$, $\boldsymbol{\beta}$ ($\in \mathbb{R}^{m(p-1)}, \mathbb{R}^{q}$ respectively), by defining $\boldsymbol{\beta}^* \in \mathbb{R}^{p}$, average of random effects coefficients

$$\boldsymbol{\beta}^* = \frac{\sum_{i=1}^{m} \mathbf{b}'_i}{m}. \tag{2.12}$$

In this way $\boldsymbol{\beta}$ is just the concatenation of $\boldsymbol{\beta}'$ and $\boldsymbol{\beta}^*$ while for $\mathbf{b}$ holds the following:

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_{m-1} \end{bmatrix} = \begin{bmatrix} \mathbf{b}'_1 - \boldsymbol{\beta}^* \\ \mathbf{b}'_2 - \boldsymbol{\beta}^* \\ \vdots \\ \mathbf{b}'_{m-1} - \boldsymbol{\beta}^* \end{bmatrix}. \tag{2.13}$$

2.5. Estimation problem

Similarly to what happens with every SR-PDE model, we estimate the unknown quantities of the model by solving, with the necessary approximations, the following minimization problem:

$$\text{Find} \underset{\beta', b', f_1, \dots, f_m}{\arg\min} J_{\Omega_i,\lambda}(\beta', b', f_1, \dots, f_m) \tag{2.14}$$

where the functional $J_{\Omega_i,\lambda}$ is defined as

$$J_{\Omega_i,\lambda}(\beta', b', f_1, \dots, f_m) = \|z - W'\beta' - V'b' - f_N\|^2 + \lambda \sum_{i=1}^{m} \int_{\Omega_i} \Delta f_i(p)^2 \, d\Omega_i, \tag{2.15}$$

or equivalently, in a more lengthy expression:

$$J_{\Omega_i,\lambda}\left(\beta', b'_1, \dots, b'_m, f_1, \dots, f_m\right) = \\ \sum_{i=1}^{m} \left(\sum_{j=1}^{n_i} \left(z_{ij} - {w'_{ij}}^T \beta' - {v'_{ij}}^T b'_i - f_i(p_{ij}) \right)^2 + \lambda \int_{\Omega_i} \Delta f_i(p)^2 \, d\Omega_i \right). \tag{2.16}$$

For simplicity, we have considered the Laplacian as differential operator, but more general choices are possible.

It shall be noticed that the unknown field must not satisfy the differential equation but contributes to the functional with the square of its misfit from the equation itself — besides its contribution in terms of distance from observed data.

We now introduce the notation necessary to tackle problem 2.14, and characterize the solution defined on suitable spaces. Assuming W'_i and V'_i for $i = 1, \dots, m$ full rank, we define the following matrices:

$$X = \begin{bmatrix} W'_1 & V'_1 & 0 & \dots & 0 & 0 \\ W'_2 & 0 & V'_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ W'_{m-1} & 0 & 0 & \dots & V'_{m-1} & 0 \\ W'_m & 0 & 0 & \dots & 0 & V'_m \end{bmatrix} \tag{2.17}$$

$$H = X\left(X^T X\right)^{-1} X^T \tag{2.18}$$

$$Q = I_N - H \tag{2.19}$$

with $X \in \mathbb{R}^{N \times (mp+q-p)}$, H and $Q \in \mathbb{R}^{N \times N}$ are the matrices that project a vector, respectively, onto the subspace spanned by the columns of X and onto its orthogonal complement with respect to $\mathbb{R}^N$. Notice that despite the matrix $X^T X$ exhibits the pattern in figure 2.1a, where only the first row, the first column and the diagonal are different from 0, the inverse of this type of matrix is in general dense, *c.f.* figure 2.1b.

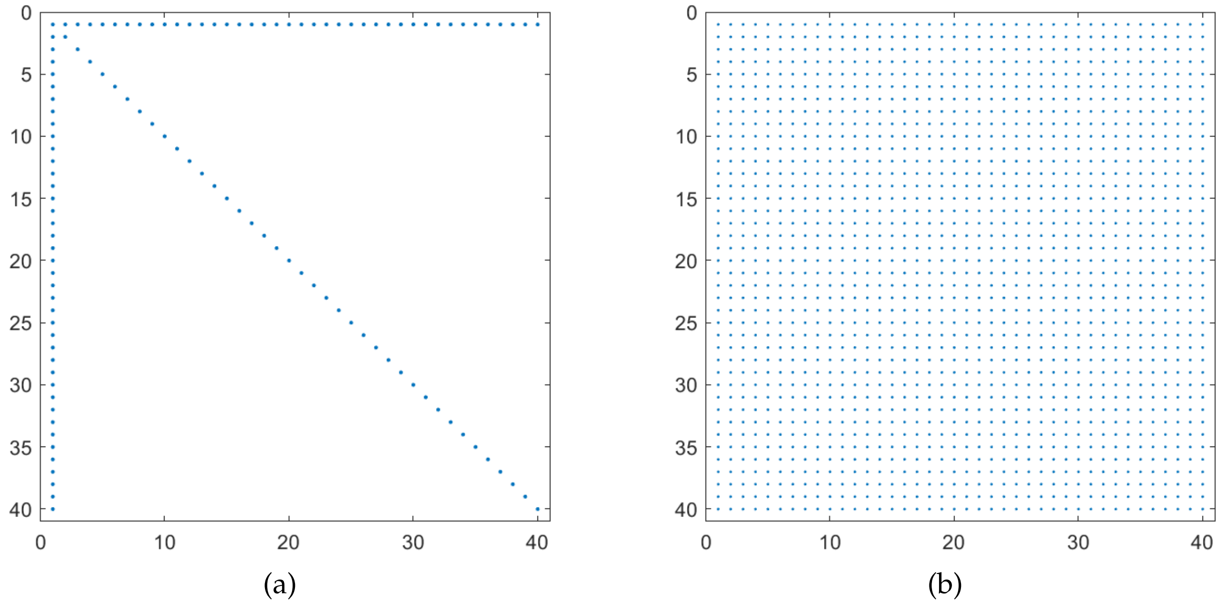

Figure 2.1.: *(a) shows the pattern of generic symmetric matrix* $X^T X$ *as defined in equation 2.17. Size has been arbitrarily chosen to be 40. On the right we show the pattern of the inverse of such matrix, illustrating that the inverse is dense. Non-zero values were sampled from a uniform distribution.*

Define also the vector of coefficients $\boldsymbol{\nu} = (\boldsymbol{\beta}', \mathbf{b}'_1, \ldots, \mathbf{b}'_m) \in \mathbb{R}^{mp+q-p}$.
Given these definitions, we can write the mixed-effects model in *implementation* version with the formula

$$z = X\boldsymbol{\nu} + \mathbf{f}_N + \boldsymbol{\epsilon}, \tag{2.20}$$

which separates the two components of our model, the parametric one and the non-parametric. The estimation functional 2.15 can then be expressed as

$$J_{\Omega_i,\lambda}(\boldsymbol{\nu}, f_1, \ldots, f_m) = \|z - X\boldsymbol{\nu} - \mathbf{f}_N\|^2 + \lambda \sum_{i=1}^{m} \int_{\Omega_i} \Delta f_i(\mathbf{p})^2 \, d\Omega_i, \tag{2.21}$$

Next section deals with the characterization of a possible minimizer $\left(\hat{\boldsymbol{\nu}}, \hat{f}_1, \ldots, \hat{f}_m\right)$ of functional 2.21.

2.6. Solving the estimation problem

Assuming that for each patient i, spatial field f_i belongs to $\mathcal{H}^2(\Omega_i)$, Sobolev space of functions whose first and second derivative are in $\mathcal{L}^2(\Omega_i)$, the functional 2.21 is well-defined.
We also make the assumption, not strictly necessary, of imposing on every field f_i homogeneous boundary condition of Neumann type, meaning the normal derivative

on the boundary is null almost everywhere. Interesting applications of different type of boundary conditions in the context of SR-PDE models are treated in [Azz+14].

The field $f = (f_1 \dots f_m)$ is therefore naturally set in the Hilbert space

$$\mathcal{V} = \bigoplus_{i=1}^{m} \mathcal{H}^2_{n0}(\Omega_i) \tag{2.22}$$

where $\mathcal{H}^2_{n0}(\Omega_i)$ is the set $\left\{g \in \mathcal{H}^2(\Omega_i) \mid \nabla g \cdot \mathbf{n} = 0 \text{ a.e. on } \partial\Omega_i\right\}$.

For the estimation problem introduced before, it holds the following:

Theorem 2.1. *With $\mathbf{v} \in \mathbb{R}^{mp+q-p}$ and $f = (f_1 \dots f_m) \in \mathcal{V}$, for functional $J_{\Omega_i,\lambda}(\mathbf{v}, f)$ there exists one and only one minimizer $\left(\hat{\mathbf{v}}, \hat{f}\right)$.*

By denoting also with $\boldsymbol{\varphi}_N$ the vector obtained by concatenation of the evaluations of a generic element φ belonging to functional space $\mathcal{V}$, at locations n_i for every i, we have the following characterization for the solution $\left(\hat{\mathbf{v}}, \hat{f}\right) = \left(\hat{\mathbf{v}}, \hat{f}_1, \dots, \hat{f}_m\right)$ of the estimation problem:

$$\begin{cases} \boldsymbol{\varphi}_N^T Q \hat{\mathbf{f}}_N + \lambda \sum_{i=1}^m \int_{\Omega_i} \Delta\varphi_i \Delta\hat{f}_i \, d\Omega_i = \boldsymbol{\varphi}_N^T Q \boldsymbol{z} \quad \forall \varphi \in \mathcal{V} \\ \hat{\mathbf{v}} = \left(X^T X\right)^{-1} X^T (\boldsymbol{z} - \hat{\mathbf{f}}_N) \end{cases} . \tag{2.23}$$

First of previous equations can be equivalently rewritten into two equations. With the introduction of an auxiliary function $\hat{g}$, whose vector components are $-\Delta\hat{f}_i$, we write system 2.23 as

$$\begin{cases} \boldsymbol{\varphi}_N^T Q \hat{\mathbf{f}}_N + \lambda \sum_{i=1}^m \int_{\Omega_i} \nabla\varphi_i \nabla g_i \, d\Omega_i = \boldsymbol{\varphi}_N^T Q \boldsymbol{z} \\ \sum_{i=1}^m \left(\int_{\Omega_i} \eta_i \hat{g}_i \, d\Omega_i + \int_{\Omega_i} \nabla\eta_i \cdot \nabla\hat{f}_i \, d\Omega_i \right) = 0 \\ \hat{\mathbf{v}} = \left(X^T X\right)^{-1} X^T (\boldsymbol{z} - \hat{\mathbf{f}}_N) \end{cases} , \tag{2.24}$$

valid for every couple (φ, η) belonging to appropriate functional spaces (essentially, we can work here with Sobolev spaces of degree one).

This form of the problem is the most suitable for the use of the finite element method to compute the desired estimation.

2.7. Numerical solution

In the library `fdaPDE` we solve the estimation problem by mean of finite elements (for detailed description see e.g. [Qua23]). We generate a triangular mesh, by partitioning domains Ω_i with a regular triangulation. Domain Ω_i is therefore approximated as $\Omega_i^{\tau_i}$, union of the triangles.

On the approximated domains we define Sobolev spaces, $\mathcal{H}^1_i\left(\Omega_i^{\tau_i}\right)$, for every i, and relative subspaces, $V^1_i\left(\Omega_i^{\tau_i}\right)$, *i.e.* the finite element space of degree one on triangulation τ_i: in this work we will limit ourselves to Lagrangian finite elements of degree one with nodes in the vertices of the triangles.

Having defined this spaces, we move to the discrete counterpart of problem 2.24. The equations remain the same if not for the domain Ω_i , approximated as $\Omega_i^{\tau_i}$. But as we change functional spaces from infinite dimension to finite dimension:

- we can write functions as linear (finite) combinations of finite element basis; we denote as $\boldsymbol{\psi}_i$ or $\boldsymbol{\psi}^{(i)}$ if necessary, the vector of fields that constitute a basis for $V_i^1\left(\Omega_i^{\tau_i}\right)$;
- we can compute integrals by numerical techniques, see [Qua23] for technical details;
- thanks to previous two points, problem 2.24 is equivalently written in algebraic form as a linear system of equations.

The following matrices definitions are used:

$$\Psi_i = \begin{bmatrix} \psi_1^{(i)}(\mathbf{p}_{i1}) & \dots & \psi_{N_{\tau_i}}^{(i)}(\mathbf{p}_{i1}) \\ \vdots & \ddots & \vdots \\ \psi_1^{(i)}(\mathbf{p}_{in_i}) & \dots & \psi_{N_{\tau_i}}^{(i)}(\mathbf{p}_{in_i}) \end{bmatrix}. \tag{2.25}$$

Ψ_i belongs to $\mathbb{R}^{n_i \times N_{\tau_i}}$, where we define N_{τ_i} as the number of basis functions deriving from triangulation τ_i relative to Ω_i. Given Neumann boundary conditions, this is also the number of nodes of the i-th mesh.

$$R_{0i} = \int_{\Omega_{\tau_i}} \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T d\Omega_{\tau_i}. \tag{2.26}$$

R_{0i} lies in $\mathbb{R}^{N_{\tau_i} \times N_{\tau_i}}$.

$$R_{1i} = \int_{\Omega_{\tau_i}} \left(\nabla \boldsymbol{\psi}_i\right)^T \nabla \boldsymbol{\psi}_i d\Omega_{\tau_i}. \tag{2.27}$$

R_{1i} lies in $\mathbb{R}^{N_{\tau_i} \times N_{\tau_i}}$ as well, with $\nabla \boldsymbol{\psi}_i$ being the matrix whose element in m-th row and n-th column is $\frac{\partial (\boldsymbol{\psi}_i)_n}{\partial x_m}$.

Before we used the operator mapping a generic element φ, belonging to functional space $\mathcal{V}$, to $\boldsymbol{\varphi}_N$, the vector obtained by concatenation of the evaluations of φ, at locations n_i for every i. By defining $\tilde{\Psi}$ as

$$\tilde{\Psi} = \begin{bmatrix} \Psi_1 & & 0 \\ & \ddots & \\ 0 & & \Psi_m \end{bmatrix}, \tag{2.28}$$

the following identity shows the role of Ψ_i matrices:

$$\boldsymbol{\varphi}_N = \tilde{\Psi} \boldsymbol{\varphi} \tag{2.29}$$

where $\boldsymbol{\varphi}$ is the coefficient vector of expansion of φ with respect to finite element basis.

Analogously as $\tilde{\Psi}$, we define the tensorised versions of R_0 and R_1 matrices as:

$$\tilde{R_0} = \begin{bmatrix} R_{01} & & 0 \\ & \ddots & \\ 0 & & R_{0m} \end{bmatrix}, \tag{2.30}$$

$$\tilde{R_1} = \begin{bmatrix} R_{11} & & 0 \\ & \ddots & \\ 0 & & R_{1m} \end{bmatrix}. \tag{2.31}$$

By mean of simple algebraic manipulations, we express equivalently the discrete counterpart of problem 2.24 with the following linear system of equations:

$$\begin{bmatrix} \tilde{\Psi}^T Q \tilde{\Psi} & -\lambda \tilde{R}_1^T \\ -\lambda \tilde{R}_1 & -\lambda \tilde{R}_0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \tilde{\Psi}^T Q \mathbf{z} \\ 0 \end{bmatrix} \tag{2.32}$$

together with the least square equation for parameter $\boldsymbol{\nu}$, which can be rewritten as

$$\hat{\boldsymbol{\nu}} = \left(X^T X \right)^{-1} X^T (\mathbf{z} - \tilde{\Psi} \hat{\mathbf{f}}) \tag{2.33}$$

We call system 2.32 monolithic because it might be tough to solve it by standard numerical techniques. In fact number of patients m might be large, and so the number of nodes N_{τ_i} for each patient i.

Therefore, aim of this work is to avoid the solution of such high dimension linear system, in favour of splitting it into many systems ($\simeq m$) of lower dimension.

In case the dimensions of the monolithic system are treatable, the Woodbury decomposition formula, described in appendix A.1, can be used to speed up the computation of the solution for different values of λ. The decomposition for system 2.32 is analogous to the one described in the appendix, with

$$E = \begin{bmatrix} \tilde{\Psi}^T \tilde{\Psi} & -\lambda \tilde{R}_1^T \\ -\lambda \tilde{R}_1 & -\lambda \tilde{R}_0 \end{bmatrix} \quad U = \begin{bmatrix} \tilde{\Psi}^T X \\ 0 \end{bmatrix} \tag{2.34}$$
$$C = -\left[(X^T X)^{-1} \right] \quad V = U^T i.$$

The general setting above is simplified in the rest of this work by assuming:

- same domain for all the patients, *i.e.* $\Omega_1 = \ldots = \Omega_m := \Omega$;
- same locations $\mathbf{p}_{ij}$ for every patient i, that is $\mathbf{p}_{ij} = \mathbf{p}_{kj} \quad \forall (i,k) \in \{1 \ldots m\}^2$. This also implies that n_i, number of observations for patient i, is the same for every i. In the following we will use $n = n_i$.
- same finite element basis used across all m domains Ω.

These properties allow storing in memory some of the matrices described above just for one patient rather than for all patients.

3. Solving large linear systems for the mixed-effects model

3.1. Iterative methods

Following the ideas stemmed from the spatio-temporal regression in Pollini and Ponti [PP13] and Massardi and Spaziani [MS21], we consider an iterative scheme. At each step i the scheme computes an approximate solution $(\hat{\mathbf{f}}^i, \hat{\mathbf{g}}^i)$ of the monolithic system by solving a single-unit problem for each statistical unit. The algorithm stops with two possible criteria:

- A maximum number of iterations is reached;

- The following two conditions are true.
  The first one is that the functional 2.15 evaluated in the estimated solution $J^i = J_{\Omega_i,\lambda}\left(\hat{\boldsymbol{\beta}}'^i, \hat{\mathbf{b}}_1'^i, \ldots, \hat{\mathbf{b}}_m'^i, \hat{f}_1^i, \ldots, \hat{f}_m^i\right)$ has reached stagnation, that is the relative increment $\left(J^i - J^{i-1}\right)/J^i$ is below a certain threshold. In the code such threshold is an input parameter, it was set to 10^{-8}. The term $\int_{\Omega_i} \Delta f_i(\mathbf{p})^2 \, d\Omega_i$ of the functional 2.15 is easily computed by exploiting the expansion into the finite element basis functions: it is equal to $\hat{\mathbf{g}}_i^T R_0 \hat{\mathbf{g}}_i$, where $\hat{\mathbf{g}}_i$ is the sub-vector of $\hat{\mathbf{g}}$ corresponding to unit i.
  The second one is that the estimated solution is close to the exact solution of the system 2.32. This condition is verified by checking that the residual, normalized by the Euclidean norm of the right-hand side, is below a certain threshold (this is another input parameter, 10^{-8} was used).

The details of the method that was implemented are described in the following sections.

3.1.1. The block diagonal approximation

Looking at the monolithic equation 2.32, the question of how to formulate an approximation of the term $\tilde{\Psi}^T Q \tilde{\Psi}$ naturally arises. In particular, a suitable block approximation allows to make the estimated field $\hat{f}_i$ of statistical unit i independent of the observations in the other units, for every unit i ($i = 1 \ldots m$). To this purpose, the following approximation is considered:

$$\tilde{\Psi}^T Q \tilde{\Psi} \simeq \Gamma := \begin{bmatrix} \Psi^T Q_1 \Psi & 0 & \dots & 0 \\ 0 & \Psi^T Q_2 \Psi & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \Psi^T Q_m \Psi \end{bmatrix}, \tag{3.1}$$

where Q_i indicates the i-th diagonal block of Q, of dimensions $n \times n$. By definition of Q, this is equal to $I_n - X_i \left(X^T X\right)^{-1} X_i^T$, with $X_i = X\left((i-1)\,n + 1 : in\,,\,:\right)$, where the typical notation of MATLAB language has been used to express a suitable submatrix. This choice is not casual but it naturally stems from the idea behind the iterative method as we will see in section 3.1.3. Not only: $\Psi^T Q_i \Psi$ is also the i-th diagonal block of matrix $\tilde{\Psi}^T Q \tilde{\Psi}$. We will show it, for example, by means of Woodbury decomposition (appendix A.1) as used in 2.34.

Express $\tilde{\Psi}^T Q \tilde{\Psi}$ as $\tilde{\Psi}^T \tilde{\Psi} - \tilde{\Psi}^T H \tilde{\Psi}$. Thus, defining $U = \tilde{\Psi}^T X$, $C = -\left(X^T X\right)^{-1}$, $V = U^T$, its i-th diagonal block is equal to $\Psi^T \Psi + U_i C V_i$, where

$$\begin{aligned} U_i = U\left((i-1)\,N_{\mathcal{T}} + 1 : iN_{\mathcal{T}}\,,\,:\right) = \tilde{\Psi}_i^T X = \\ \begin{bmatrix} 0 & \dots & 0 & \Psi^T & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_m \end{bmatrix} = \Psi^T X_i, \end{aligned} \tag{3.2}$$

$$V_i = U_i^T = X_i^T \Psi. \tag{3.3}$$

Thus, obtaining what we wanted to prove,

$$\Psi^T \Psi + U_i C V_i = \Psi^T \left(I_n - X_i \left(X^T X\right)^{-1} X_i^T \right) \Psi = \Psi^T Q_i \Psi. \tag{3.4}$$

3.1.2. Initialization

The initialization consists in finding a good guess $(\hat{\mathbf{f}}^0, \hat{\mathbf{g}}^0)$ to start the algorithm from; for this purpose, the following m problems are solved: for $i = 1, \dots, m$ solve

$$\begin{bmatrix} \Psi^T Q_i \Psi & -\lambda R_1^T \\ -\lambda R_1 & -\lambda R_0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}}_i^0 \\ \hat{\mathbf{g}}_i^0 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_i \\ 0 \end{bmatrix}, \tag{3.5}$$

where $\mathbf{u}_i$ is the vector whose components are the first n components of $\tilde{\Psi}^T Q \mathbf{z}$ starting from the $n(i-1) + 1$-th component.

3.1.3. Iterations

The idea behind the iterative scheme is, having a guess of a solution $(\hat{\mathbf{f}}^{k-1}, \hat{\mathbf{g}}^{k-1})$, to compute a new guess $(\hat{\mathbf{f}}^k, \hat{\mathbf{g}}^k)$ by replacing inside the monolithic system 2.32 the unknown coefficients relative to all but one unit with their value computed at previous step, for every unit.

For example, the first n equations of the monolithic system 2.32 read

$$\begin{bmatrix}\Psi^T Q_{1,1}\Psi & \Psi^T Q_{1,2}\Psi & \dots & \Psi^T Q_{1,m}\Psi & -\lambda R_1^T\end{bmatrix}\begin{bmatrix}\hat{\mathbf{f}}_1\\ \hat{\mathbf{f}}_2\\ \vdots\\ \hat{\mathbf{f}}_m\\ \hat{\mathbf{g}}_1\end{bmatrix} = \begin{bmatrix}\mathbf{u}_1\end{bmatrix}, \tag{3.6}$$

where $Q_{i,j}$ indicates block of row i and column j of matrix Q. Substituting $\hat{\mathbf{f}}_j$ for $j \neq 1$ with $\hat{\mathbf{f}}_j^{k-1}$, taking previous step values to the right-hand side and generalizing, the iterative scheme reads: for $i = 1, \dots, m$ solve

$$\begin{bmatrix}\Psi^T Q_i \Psi & -\lambda R_1^T\\ -\lambda R_1 & -\lambda R_0\end{bmatrix}\begin{bmatrix}\hat{\mathbf{f}}_i^k\\ \hat{\mathbf{g}}_i^k\end{bmatrix} = \begin{bmatrix}\mathbf{r}_i\\ 0\end{bmatrix} \tag{3.7}$$

where

$$\mathbf{r}_i = \mathbf{u}_i - \sum_{\substack{j=1\\ j\neq i}}^{m} \Psi^T Q_{i,j}\Psi\hat{\mathbf{f}}_j^{k-1} \tag{3.8}$$

An estimate of $\mathbf{v}$, $\hat{\mathbf{v}}^i = (\hat{\boldsymbol{\beta}}'^i, \hat{\mathbf{b}}_1'^i, \dots, \hat{\mathbf{b}}_m'^i)$ has to be computed at each iteration for the estimation of the functional 2.15, according to equation 2.33.

3.1.4. The iterative method as a preconditioned Richardson method

Given a generic preconditioning matrix P, calling $\mathbf{r}_k$ the residual of the linear system $A\mathbf{x} = \mathbf{b}$ (that is the vector $A\mathbf{x}_k - \mathbf{b}$) at step k, Richardson method (*cf.*, for example, [Ger+14])consists in solving (or rather trying to) the linear system iterating the following steps:

1. Solve $P\mathbf{z}_k = \mathbf{r}_k$
2. Compute the acceleration parameter α_k (for simplicity we use $\alpha_k = 1$)
3. Update the solution $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{z}_k$
4. Update the residual $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k A\mathbf{z}_k$

The iterative scheme described in the previous section is a Richardson scheme with the following preconditioning matrix:

$$P = \begin{bmatrix}\Gamma & -\lambda\tilde{R}_1^T\\ -\lambda\tilde{R}_1 & -\lambda\tilde{R}_0\end{bmatrix} \tag{3.9}$$

Solving a large linear system involving this matrix is indeed like solving m linear systems of dimensions $2n \times 2n$.

3.2. Generalized cross validation

Problem 2.32 is solved multiple times on a grid of λs. The best λ is then chosen according to a generalized cross-validation criterion. In particular, as in [San21], the minimum of the generalized cross-validation (GCV) parameter is used as a model selection criterion. The generalization of this parameter to the mixed-effect model reads as follows:

$$\mathrm{GCV}(\lambda) = \frac{1}{N\left(1 - \left(q - p + mp + \mathrm{tr}\left(S\right)\right)/N\right)^2} \|z - \hat{z}\|^2 \tag{3.10}$$

where for q, m, p we follow the notation of section 2.4, and S indicates the smoothing matrix, which is the matrix that maps the observations vector z into the estimated spatial field $\hat{\mathbf{f}}$ evaluated at the location of the observations ($\hat{\mathbf{f}}_N = \tilde{\Psi}\hat{\mathbf{f}} = Sz$). Its value is

$$S = \tilde{\Psi}\left(\tilde{\Psi}^T Q \tilde{\Psi} + \lambda \tilde{R}_1^T \tilde{R}_0^{-1} \tilde{R}_1\right)^{-1} \tilde{\Psi}^T Q, \tag{3.11}$$

and it stems as a Schur complement for system 2.32 with respect to the estimated field $\hat{\mathbf{f}}$.

The computation of the GCV parameter 3.10 involves the expensive computation of the trace of the smoothing matrix S. Since for any two matrices A and B such that the product AB is defined, it holds that $\mathrm{tr}\left(AB\right) = \mathrm{tr}\left(BA\right)$, the trace of S can be computed as the trace of

$$\tilde{S} = \left(\tilde{\Psi}^T Q \tilde{\Psi} + \lambda \tilde{R}_1^T \tilde{R}_0^{-1} \tilde{R}_1\right)^{-1} \tilde{\Psi}^T Q \tilde{\Psi}. \tag{3.12}$$

4. Statistical inference

4.1. Properties of the estimators

4.1.1. Spatial field component

To study the statistical properties of the estimated spatial field that is obtained by solving the monolithic system 2.32, we write it in the following equivalent manner

$$\begin{cases} \tilde{\Psi}^{T} Q \tilde{\Psi} \hat{\mathbf{f}} - \lambda \tilde{R}_1^{T} \hat{\mathbf{g}} = \tilde{\Psi}^{T} Q \mathbf{z} \\ \hat{\mathbf{g}} = -\tilde{R}_0^{-1} \tilde{R}_1 \hat{\mathbf{f}} \end{cases} \tag{4.1}$$

By recalling the mixed-effect model in the implementative form 2.20, that is $z = X\boldsymbol{\nu} + \mathbf{f}_N + \boldsymbol{\epsilon}$, and substituting the second equation inside the first we get

$$\tilde{\Psi}^{T} Q \tilde{\Psi} \hat{\mathbf{f}} + \lambda \tilde{R}_1^{T} \tilde{R}_0^{-1} \tilde{R}_1 \hat{\mathbf{f}} = \tilde{\Psi}^{T} Q \left(X\boldsymbol{\nu} + \mathbf{f}_N + \boldsymbol{\epsilon} \right) \tag{4.2}$$

Recalling that Q projects on the orthogonal space of the columns of X, implying $QX = 0$,

$$\left(\tilde{\Psi}^{T} Q \tilde{\Psi} + \lambda \tilde{R}_1^{T} \tilde{R}_0^{-1} \tilde{R}_1 \right) \hat{\mathbf{f}} = \tilde{\Psi}^{T} Q \mathbf{f}_N + \tilde{\Psi}^{T} Q \boldsymbol{\epsilon} \tag{4.3}$$

That can be rearranged as follows

$$\left(\tilde{\Psi}^{T} Q \tilde{\Psi} + \lambda \tilde{R}_1^{T} \tilde{R}_0^{-1} \tilde{R}_1 \right) \left(\hat{\mathbf{f}} - \mathbf{f} \right) + \lambda \tilde{R}_1^{T} \tilde{R}_0^{-1} \tilde{R}_1 \mathbf{f} = \tilde{\Psi}^{T} Q \boldsymbol{\epsilon} \tag{4.4}$$

where we have defined $\mathbf{f}$ as the coefficients of the finite element expansion of the true vector field $f(\cdot)$. Deriving now the term $\hat{\mathbf{f}} - \mathbf{f}$

$$\hat{\mathbf{f}} - \mathbf{f} = -\left(\tilde{\Psi}^{T} Q \tilde{\Psi} + \lambda \tilde{R}_1^{T} \tilde{R}_0^{-1} \tilde{R}_1 \right)^{-1} \lambda \tilde{R}_1^{T} \tilde{R}_0^{-1} \tilde{R}_1 \mathbf{f} + \left(\tilde{\Psi}^{T} Q \tilde{\Psi} + \lambda \tilde{R}_1^{T} \tilde{R}_0^{-1} \tilde{R}_1 \right)^{-1} \tilde{\Psi}^{T} Q \boldsymbol{\epsilon} \tag{4.5}$$

The first term above is not random, thus we can derive the following:

1. Thanks to the property that, for generic suitably dimensioned matrices and vectors, $\mathbb{E}[M\mathbf{x}] = M\mathbb{E}[\mathbf{x}]$, we get that $\mathbb{E}\left[\hat{\mathbf{f}} - \mathbf{f}\right] = -\left(\tilde{\Psi}^{T} Q \tilde{\Psi} + \lambda \tilde{R}_1^{T} \tilde{R}_0^{-1} \tilde{R}_1 \right)^{-1} \lambda \tilde{R}_1^{T} \tilde{R}_0^{-1} \tilde{R}_1 \mathbf{f}$;

2. Thanks to the property that again, for generic suitably dimensioned matrices and vectors, $Var[M\mathbf{x}] = M Var[\mathbf{x}] M^{T}$, we get this result for $Var\left[\hat{\mathbf{f}} - \mathbf{f}\right] = Var\left[\hat{\mathbf{f}}\right]$

$$Var\left[\hat{\mathbf{f}} - \mathbf{f}\right] = \sigma^2 \left(\tilde{\Psi}^{T} Q \tilde{\Psi} + \lambda \tilde{R}_1^{T} \tilde{R}_0^{-1} \tilde{R}_1 \right)^{-1} \tilde{\Psi}^{T} Q \tilde{\Psi} \left(\tilde{\Psi}^{T} Q \tilde{\Psi} + \lambda \tilde{R}_1^{T} \tilde{R}_0^{-1} \tilde{R}_1 \right)^{-1} \tag{4.6}$$

where we have exploited symmetric matrices and the fact that $Q^2 = Q$.

4.1.2. Parametric component

For what concerns the estimation of the parameter $\boldsymbol{\nu}$, we recall that $\hat{\boldsymbol{\nu}}$ solves the normal equations

$$X^T X \hat{\boldsymbol{\nu}} = X^T \left(\boldsymbol{z} - \tilde{\Psi} \hat{\mathbf{f}} \right) \tag{4.7}$$

that with model 2.20 can be written as

$$X^T X \hat{\boldsymbol{\nu}} = X^T \left(X\boldsymbol{\nu} + \tilde{\Psi}\mathbf{f} + \boldsymbol{\epsilon} - \tilde{\Psi} \hat{\mathbf{f}} \right) \tag{4.8}$$

or

$$X^T X \left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu} \right) = -X^T \tilde{\Psi} \left(\hat{\mathbf{f}} - \mathbf{f} \right) + X^T \boldsymbol{\epsilon} \tag{4.9}$$

From this we can derive:

$$\mathbb{E}\left[\hat{\boldsymbol{\nu}} - \boldsymbol{\nu} \right] = \left(X^T X \right)^{-1} X^T \tilde{\Psi} \left(\tilde{\Psi}^T Q \tilde{\Psi} + \lambda \tilde{R}_1^T \tilde{R}_0^{-1} \tilde{R}_1 \right)^{-1} \lambda \tilde{R}_1^T \tilde{R}_0^{-1} \tilde{R}_1 \mathbf{f} \tag{4.10}$$

and the expression of $Var\left[\hat{\boldsymbol{\nu}} - \boldsymbol{\nu} \right]$ that is not reported for brevity.

4.1.3. Asymptotic properties

We study the limiting properties of the estimators in the setting where the number of basis $N_{\mathcal{T}}$ and the triangulation $\mathcal{T}$ are fixed. The number of observation n_i for each statistical unit increases to infinity.

In a similar fashion as in [SFF21], where the simple spatial problem is considered, we make some assumptions for studying the asymptotic properties of the considered estimators.

5. Simulation studies

In this chapter we see some examples...

6. Applications

aa

7. Conclusions

Aaa

Appendices

A. Formulae

A.1. Woodbury decomposition

The following matrix identity holds.

Proposition A.1 (Woodbury matrix identity)**.** *Let* M *be a square* $m \times m$ *matrix which can be written as the sum* $E + UCV$*, with E being* $m \times m$*,* U *being* $m \times n$*,* C *being a square* $n \times n$ *matrix, and* V $n \times m$*. Then*

$$M^{-1} = (E + UCV)^{-1} = E^{-1} - E^{-1}U\left(C^{-1} + VE^{-1}U\right)^{-1} VE^{-1} \tag{A.1}$$

In particular, in the case where inverting matrix E can be considered cheap or useful, on the righ-hand side solving the system involves solving an $n \times n$ system (the one carachterized by matrix $C^{-1} + VE^{-1}U$), whilst on the left the dimensions are $m \times m$. This equation is exploited for faster system solving in the `fdaPDE` library. As an example, consider the space-only problem, described *e.g.* in [San21]: the presence of covariates leads to a linear system involving the following matrix M:

$$M = \begin{bmatrix} \Psi^T Q \Psi & -\lambda R_1^T \\ -\lambda R_1 & -\lambda R_0 \end{bmatrix}$$

Remembering that the projection matrix Q is defined as $I - H$, M can be split into the following two components, one independent from λ:

$$M = \begin{bmatrix} \Psi^T \Psi & -\lambda R_1^T \\ -\lambda R_1 & -\lambda R_0 \end{bmatrix} + \begin{bmatrix} -\Psi^T H \Psi & 0 \\ 0 & 0 \end{bmatrix}$$

By defining E the left matrix of the two above (which is also the matrix corresponding to the problem without covariates), and remembering that $H = W\left(W^T W\right)^{-1} W^T$, Woodbury decomposition can be exploited defining the following matrices U, C, V:

$$\begin{bmatrix} -\Psi^T W\left(W^T W\right)^{-1} W^T \Psi & 0 \\ 0 & 0 \end{bmatrix} = UCV$$

$$U = \begin{bmatrix} \Psi^T W \\ 0 \end{bmatrix}$$

$$C = -\left[\left(W^T W\right)^{-1}\right]$$

$$V = U^T = \begin{bmatrix} W^T\Psi & 0 \end{bmatrix}$$

where the 0s indicate matrices of zeros of suitable dimensions.
Since matrices U, C, V do not depend on λ, computing the solution of the system for different values of λ involves only the factorization of the matrix E (and the cheaper $C^{-1} + VE^{-1}U$).

Bibliography

[Azz+14] L. Azzimonti, F. Nobile, L. M. Sangalli, and P. Secchi. "Mixed Finite Elements for Spatial Regression with PDE Penalization". In: *SIAM/ASA Journal on Uncertainty Quantification* 2.1 (2014), pp. 305–335. DOI: `10.1137/130925426`. URL: `https://doi.org/10.1137/130925426` (cit. on p. 21).

[BHT89] A. Buja, T. Hastie, and R. Tibshirani. "Linear Smoothers and Additive Models". In: *The Annals of Statistics* 17 (June 1989). DOI: `10.1214/aos/1176347115`.

[Ger+14] P. Gervasio, A. Quarteroni, R. Sacco, and F. Saleri. *Matematica Numerica*. Feb. 2014. ISBN: 978-88-470-5644-2 (cit. on p. 27).

[Kim20] J. Kim. "Mixed-Effect Models in Spatial Regression with Partial Differential Equation Regularization". Politecnico di Milano, 2020 (cit. on pp. 16, 17).

[MS21] M. Massardi and S. Spaziani. *fdaPDE: an efficient iterative method for spatio-temporal regression with PDE regularization*. Politecnico di Milano, APSC course, 2021 (cit. on p. 25).

[PP13] M. Pollini and L. Ponti. *Regressione spazio-temporale su domini bidimensionali non planari*. Politecnico di Milano, NAPDE course, 2013 (cit. on p. 25).

[Qua23] A. Quarteroni. "Modellistica Numerica per Problemi Differenziali (Third. Ed.)" In: (Apr. 2023) (cit. on pp. 21, 22).

[San21] L. Sangalli. "Spatial Regression With Partial Differential Equation Regularisation". In: *International Statistical Review* (Mar. 2021). DOI: `10.1111/insr.12444` (cit. on pp. 15, 28, 39).

[SFF21] L. Sangalli, F. Ferraccioli, and L. Finos. "Some first inferential tools for spatial regression with differential regularization". In: *Journal of Multivariate Analysis* (2021) (cit. on p. 30).

[SRR13] L. Sangalli, J. Ramsay, and T. Ramsay. "Spatial spline regression models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75 (Sept. 2013). DOI: `10.1111/rssb.12009` (cit. on p. 15).