DEPARTMENT OF
MATHEMATICS

STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS AND APPLICATIONS IN BAYESIAN INFERENCE

SÍLVIA ISABEL BELO GUERRA
Research Master in Finance
Master in Mathematics and Applications

DOCTORATE IN STATISTICS AND RISK MANAGEMENT

NOVA University Lisbon
May, 2022

DEPARTMENT OF
MATHEMATICS

STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS AND APPLICATIONS IN BAYESIAN INFERENCE

SÍLVIA ISABEL BELO GUERRA

Research Master in Finance
Master in Mathematics and Applications

Adviser: Isabel Cristina Maciel Natário
Associate Professor, NOVA University Lisbon

Co-adviser: Maria Fernanda de Almeida Cipriano Salvador Marques
Associate Professor, NOVA University Lisbon

Examination Committee

Chair: Carlos Manuel Agra Coelho
Full Professor, Nova School of Science and Technology - FCT NOVA

Rapporteurs: Nikolai Vasilievich Chemetov
Full Professor, Computer Science Department - University of São Paulo

Gonçalo José Nunes dos Reis
Associate Professor, School of Mathematics - University of Edinburgh

Co-adviser: Maria Fernanda de Almeida Cipriano Salvador Marques
Associate Professor, Nova School of Science and Technology - FCT NOVA

Members: Patrícia Cortés de Zea Bermudez
Assistant Professor, Faculty of Sciences - University of Lisbon

Carlos Manuel Agra Coelho
Full Professor, Nova School of Science and Technology - FCT NOVA

Manuel Leote Tavares Inglês Esquível
Associate Professor, Nova School of Science and Technology - FCT NOVA

DOCTORATE IN STATISTICS AND RISK MANAGEMENT

NOVA University Lisbon
May, 2022

Stochastic Partial Differential Equations and Applications in Bayesian Inference

This document was created with the (pdf/Xe/Lua)LaTeX processor and the NOVAthesis template (v6.9.6) [30].

To my mother, to my father, to my husband and my children.

Acknowledgements

I would like to express my most sincere gratitude to my advisor, Professor Isabel Natário, and my co-adviser, Professor Fernanda Cipriano, for all the help, guidance and support. I am specially grateful for all their dedication and care. Their bright minds are for me a source of inspiration to go further.

I would like also to express my profound gratitude to Professor Manuel L. Esquível for all the help, kindness, and care.

My sincere thanks to Doctor Ricardo Tomé, from Instituto Dom Luiz, for the wind data set: the H. Hersbach et al. (2018) was downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store. The results contain modified Copernicus Climate Change Service information 2021. It should be mentioned that neither the European Commission nor the European Centre for Medium-Range Weather Forecasts is responsible for any use that may be made of the Copernicus information or data it contains.

I also thank FCT - Foundation for Science and Technology - under the projects UID/MAT/00297/2019 and PTDC/MAT-STA/28243/2017, for financing my participation in the conference Statistical Analysis for Space-Time Data (ECAS2019), and in NOVA Science Day 2019.

I would like to express a very special thanks to Professor Patrícia Xufre, for her support, friendship, and her special advice in a time of need.

I am deeply grateful to my mother and to my father for all the effort they've made, so I could go on with my academic studies. I also thank them for their support, patience and love. They were always a safe shelter. I also thank my sister very much for all the help when I needed most.

I am also deeply grateful to my husband for all the patience, dedication and love. His support was essential throughout this journey.

Abstract

This work is devoted to the study of stochastic partial differential equations and its applications in Bayesian inference. Essentially it is composed by two essential parts: In the first part, we study a stochastic non-linear partial differential equation from the theoretical point of view, and in the second one we perform a Bayesian analysis based on a specific stochastic linear partial differential equation.

The former problem to be addressed has its roots in fluid dynamics. More precisely, we consider the equation which governs the time evolution of a third grade non-Newtonian fluid, filling a two-dimensional non axisymmetric bounded domain, perturbed by a multiplicative white noise. We recall that the stochastic third grade fluid equation can be considered as a generalization of the stochastic Navier-Stokes equations, so we are faced with a strongly nonlinear stochastic partial differential equation and its analysis is not an easy issue. Considering initial conditions in the Sobolev space H^2, and a Navier slip boundary condition, we show the existence and the uniqueness of the strong solution (in the stochastic sense). To show the existence of the solution, we first construct a sequence of solutions for the finite dimensional approximate problems, by using the finite dimensional Galerkin approximation method. Next, we pass to the limit by using a conjugation of compactness results and a uniqueness type argument. Let us mention that the study of stochastic fluid dynamics equations, where the solutions correspond to stochastic processes defined on some probability space, with sample paths on appropriate functional spaces is crucial for the statistical description of turbulent flows. In contrast to the usual deterministic individual solutions, in this framework each solution should correspond to a collection of possible realizations, and a probability of certain occurrences should be determined. As many fluids used in the industry are classified as third grade non-Newtonian fluid, we hope that our result will have practical consequences in the analysis of turbulence flows. As far as we know, this is the first time that the stochastic third grade fluid equation is being studied in the literature.

The second problem to be studied consists on the application of the INLA methodology to perform Bayesian inference, considering a certain linear stochastic partial differential equation, which has a solution with a Matérn covariance. We recall that the

Matérn covariance has a central role in spatial statistics, since it successfully captures the spatial behaviour of a wide number of phenomena. We consider a Gaussian vector field modelling the velocity of the wind and perform a Bayesian analysis to approximate the mean of the wind velocity field through the INLA methodology, combined with stochastic partial differential equations (SPDE). We emphasize that the behaviour of the wind velocity field is crucial in the weather forecast. We expect that this new statistical method will improve the classical methods mainly based on the numerical analysis of complex fluid dynamic equations.

Keywords: Bayesian inference, Gaussian Markov random field, Matérn covariance, Non-newtonian fluid, Stochastic partial differential equation.

Resumo

O presente trabalho é dedicado ao estudo de equações diferenciais parciais estocásticas e à sua aplicação na inferência Bayesiana. É composto essencialmente por duas partes. Na primeira parte estudamos uma equação diferencial parcial estocástica não-linear do ponto de vista teórico, e na segunda parte aplicamos os princípios da inferência Bayesiana à estimação usando uma equação diferencial parcial estocástica linear.

O primeiro problema a ser estudado tem as suas origens na dinâmica dos fluidos. Mais precisamente, consideramos a equação que descreve a evolução de um fluido não-Newtoniano de terceiro grau, num domínio bi-dimensional limitado e não axissimétrico, perturbada por um ruído branco. A equação estocástica de fluidos de terceiro grau pode ser considerada como uma generalização da equação de Navier-Stokes, estamos perante uma equação diferencial parcial estocástica fortemente não linear, cuja análise é uma tarefa difícil. Considerando a condição inicial no espaço de Sobolev H^2 , e uma condição de fronteira de deslizamento do tipo Navier, mostramos a existência e unicidade de solução forte (no sentido estocástico). Para mostrar a existência de solução, construímos primeiro uma sucessão de soluções para o problema aproximado em dimensão finita, usando o método de aproximação de Galerkin. A seguir é feita a passagem ao limite, através de resultados de compacidade, e um argumento de unicidade. Referimos também que o estudo de equações estocásticas de fluidos, cujas soluções correspondem a processos estocásticos definidos num determinado espaço de probabilidade, com trajetórias em espaços funcionais apropriados, é crucial na descrição de fluxos turbulentos. No contexto estocástico, cada solução da equação corresponde a uma coleção de possíveis realizações, pelo que a probabilidade de ocorrência de certas realizações pode ser determinada. Uma vez que muitos fluidos usados na indústria são classificados como fluidos não-Newtonianos de terceiro grau, esperamos que os nossos resultados venham a ter aplicação na análise da turbulência de fluidos. Tanto quanto pudemos constatar, esta é a primeira vez que a equação estocástica de fluidos de terceiro grau é estudada na literatura.

O segundo problema a ser estudado consiste na aplicação da metodologia INLA, que é especialmente adequada para fazer inferência Bayesiana, combinada com a utilização de determinada equação diferencial parcial estocástica, cuja solução apresenta covariância de Matérn. A covariância de Matérn tem um papel central na estatística espacial, uma vez que descreve significativamente bem vários fenómenos de natureza espacial. Neste trabalho, consideramos que a velocidade do vento é modelada por um campo vetorial aleatório Gaussiano, e aproximamos a média do campo de velocidades aplicando os princípios da inferência Bayesiana, através da metodologia INLA combinada com SPDE. Salientamos que a velocidade do vento é crucial na previsão do tempo, então esperamos que esta nova abordagem estatística, venha a melhorar os métodos usuais de previsão, baseados na análise numérica das equações de fluidos.

Palavras-chave: Campo aleatório Gaussiano de Markov, Covariância de Matérn, Equação diferencial parcial estocástica, Fluido não-Newtoniano, Inferência Bayesiana.

Contents

List of Figures

List of Tables

Acronyms

FEM Finite Element Method *(pp. xv, 1, 3, 45, 56, 67, 73, 83)*

GF Gausian field *(pp. 3, 49, 84)*

GMRF Gaussian Markov random field *(pp. 1, 3, 50–52, 55, 56, 72, 73, 76, 77, 84)*

GRF Gaussian random fields *(p. 84)*

INLA integrated nested Laplace approximation *(pp. xiv, 1, 3, 56, 61, 63, 76, 77, 79, 81, 83, 84)*

SPDE stochastic partial differential equations *(pp. x, xii, 1–3, 45, 47, 49, 52–54, 67, 72, 83, 84)*

1 Introduction

The purpose of our work is twofold: to study SPDE and to do Bayesian inference in models using SPDE as their components.

First we study the stochastic third grade fluid equation from a theoretical point of view, proving the existence and uniqueness of strong solution (Chapter 3), then we expose in detail the calculations that allow us to approximate a Matérn field by a Gaussian Markov random field (GMRF): we show in detail that the solution of a certain SPDE has Matérn covariance, widely used in geostatistic, then we apply the FEM to obtain a finite representation of the solution with Markov properties (Chapter 4). Finally, we consider the H. Hersbach et al. (2018) data set ([26]), and perform Bayesian inference to model the mean of the wind velocity, via an INLA approach (Chapter 5).

We start our work with a fundamental result: the proof of existence and uniqueness of solution for the third grade fluid equation, perturbed with a multiplicative white noise,

$$d(v(Y)) = \Big(-\nabla p + \nu\Delta Y - (Y\cdot\nabla)v - \sum_j v^j \nabla Y^j + (\alpha_1+\alpha_2)\operatorname{div}\Big(A^2\Big)$$
$$+\beta\operatorname{div}\Big(|A|^2 A\Big) + U\Big)dt + \sigma(t,Y)\,d\mathscr{W}_t, \tag{1.0.1}$$

where $v(y) = y - \alpha_1\Delta y$, $A = \nabla y + \nabla y^T$, $\nu \geq 0$ is the viscosity coefficient, U represents a body force, $\mathscr{W}_t$ is a Wiener process, and $\alpha_1, \alpha_2, \beta$ are constants such that

$$\beta \geq 0, \quad \alpha_1 \geq 0, \quad |\alpha_1 + \alpha_2| \leq \sqrt{24\nu\beta}\,,$$

therefore extending the deterministic results of [10] and [8], for the stochastic case. Equation (1.0.1) is supplemented with a Navier slip boundary condition, and the initial condition is taken in the Sobolev Space H^2.

Third grade fluid equation describes a special type of fluid, characterized by a non-linear relation between the shear stress and the shear strain rate. This means that these fluids do not satisfy the Newton's law of viscosity, so they belong to the class of non-Newtonian fluids.

From a theoretical point of view, equation (1.0.1) is a strongly nonlinear partial differential equation, which models complex viscoelastic fluids, so it is expected that the noise perturbations should have relevant impact in the fluid dynamic. It is well known that increasing the typical velocity will increase the Reynolds number: the fluid develops a turbulent behaviour and small disturbances should have strong macroscopic effects on the dynamic.

It should be pointed out that non-Newtonian fluids are present in biology, industry, etc. (see [21], [19], [24], [25], [32]). Moreover, the second grade fluid model does not capture certain specific properties, for instance, the shear thinning and shear thickening effects, so it is important to study the third grade fluid model.

We should refer the pioneer work [3] on the stochastic Navier-Stokes equations, and [4] where the authors deduce the stochastic Navier-Stokes equations from fundamental principles, showing that the stochastic equations are real physical models. Regarding the stochastic description of a Newtonian fluid, see [13], [15].

We recall that the strategy to show the existence of the solution for the deterministic fluid equation in [10] and [8], is based on appropriated estimates that allow to use compactness theorems in order to pass the non linear terms to the limit, in the weak sense. For the stochastic case, it is not possible to use such a strategy, due to lack of regularity with respect to time, and to the stochastic variable. Instead, we apply the methods developed in [6]. Those methods have been successfully applied to the stochastic second grade fluid equation (see [12] and [33]). We use the finite dimensional Galerkin approximation method to construct a sequence of solutions for the finite dimensional approximate problems, and then we pass to the limit by using a conjugation of compactness results and a unicity argument.

After establishing the existence and uniqueness of strong solution for the stochastic third grade model, we turn our attention to a certain linear SPDE, with a very specific objective: to estimate the solution of this particular equation, whose properties are particularly useful for modelling spatial phenomena. For instance, consider a random variable Y, describing a spatial phenomena, such as velocity of the wind, temperature, or pressure, to which we attribute a density probability function $\pi(y|\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta$ is an unknown parameter or vector of unknown parameters. Since θ is unknown, Bayesian Inference states that we should consider it as a random variable, with a suitable *a priori* density function $\pi(\theta)$. This prior probability function incorporates all the previous knowledge about the phenomena that is being studied.

The next step is to upgrade the *a priori* probabilities, using observed data, if $\mathbf{y} = (y_1, \ldots, y_n)$ is a vector of observations we obtain the posterior density probability function $\pi(\boldsymbol{\theta}|y)$, via Bayes' Theorem,

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathbf{y})} = \frac{\pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_\Theta \pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}}\,, \quad \boldsymbol{\theta} \in \Theta\,.$$

Observations update prior probabilities and the knowledge about theta.

The models that we are interested in this work involve a SPDE, whose solution presents a Matérn covariance function. The Matérn covariance has a central role in spatial statistics, since it successfully captures the spatial behaviour of a wide number of phenomena: forestry, disease mapping, mining engineering, etc. We assume that the spatial phenomena to be studied is modelled by a Gaussian spatial process, $\{x(s), s \in \mathcal{D} \subset \mathbb{R}^d\}$, also known as Gausian field (GF), whose mean μ does not depend on the location s, so $x(s)$ is stationary, and the covariance function depends only on the Euclidean distance between two locations, so $x(s)$ is isotropic. The Matérn covariance function evolves from this setting, and was developed by Bertil Matérn (1960). It is given by

$$\mathrm{Cov}(x(s), x(s')) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (k\|s - s'\|)^\nu K_\nu(k\|s - s'\|) \,, \tag{1.0.2}$$

where K_ν is the modified Bessel function of second kind and order ν. One special feature of this covariance function, is the relation between the parameter ν and the differentiability of the underlying process (see [22]). This is an important feature because differentiability affects the behaviour of predictions made under the model. The interpretation of ν as a smoothness parameter of the underlying spatial process is often used to justify its choice. A GF with Matérn covariance function is also called Matérn field.

We consider the following SPDE,

$$(k^2 - \Delta)^{\alpha/2} \tau x = \mathcal{W} . \tag{1.0.3}$$

where $\mathcal{W}$ is the Gaussian white noise, whose stationary solution is a Matérn field (see [38] and [39]). Lindgren et al. in [29], applying the FEM to equation (1.0.3), obtain a finite representation $\mathbf{x}$ of the solution $x(s)$, with Markov properties, and they prove convergence of the finite representation to the solution $x(s)$. Their approach allows to approximate the GF by a GMRF with sparse precision matrix, which brings great computational benefits. We present and explore the results that allow us to obtain the relation between a GF with Matérn covariance, and a finite representation by a GMRF. Before going through the FEM, it is important to understand fully the notion of GMRF. Following [36], a GMRF is defined in almost a computational point a view, as a random vector that verifies certain properties with respect to a certain *labelled graph*. After that, we explain the FEM and how can we calculate the precision matrix of the GMRF approximation $\mathbf{x}$.

As far as we know the calculations which leads to the Matérn covariance formula (1.0.2) for the stationary solution, or the Laplace method for Gaussian approximations, are not fully detailed in the literature. We have part of the calculations in [5], or [36], or [39] for instance, but all the theory that allows us to solve equation (1.0.3) is not gathered in one single source, mainly due to the diversity of subjects involved, ranging from the notion of generalized functions (also known as *distributions*), to the generalization of the Fourier transform to generalized functions. Also, we need to carefully introduce the white noise, and define the convolution between a generalized random function and an element of $L^2(\mathbb{R}^d)$. This work fills the need to have all detailed calculations and concepts in one single source.

We use the INLA methodology to obtain the posterior distribution of the hyperparameters. INLA was developed by Rue et al. in [37], where the authors perform approximate

Bayesian inference for a special class of hierarchical models, *latent Gaussian models*. The marginal distributions of the posterior distributions $\pi(\boldsymbol{\theta}|\mathbf{y})$ and $\pi(\mathbf{x}|\mathbf{y})$, can be written as

$$\pi(x_i|\mathbf{y}) = \int \pi(x_i|\boldsymbol{\theta},\mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y})\,d\boldsymbol{\theta}$$
$$\pi(\theta_j|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})\,d\boldsymbol{\theta}_{-j}$$

so the authors in [37] use approximated distributions $\tilde{\pi}(x_i|\boldsymbol{\theta},\mathbf{y})$ and $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$, based on the following results,

$$\begin{aligned}\pi(\boldsymbol{\theta}|\mathbf{y}) &\propto \frac{\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})} \\ &\approx \left.\frac{\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})}\right|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})} =: \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}),\end{aligned}$$

where $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})$ is the Gaussian approximation given by the Laplace method, explained in Section 5.3, and $\mathbf{x}^*(\boldsymbol{\theta})$ is the mode of that distribution, for a given $\boldsymbol{\theta}$. Furthermore, they apply once more the Laplace method to obtain a Gaussian approximation of $\pi(x_i|\boldsymbol{\theta},\mathbf{y})$,

$$\tilde{\pi}(x_i|\boldsymbol{\theta},\mathbf{y}) \propto \left.\frac{\pi(\mathbf{x},\boldsymbol{\theta},\mathbf{y})}{\tilde{\pi}_G(\mathbf{x}_{-i}|x_i,\boldsymbol{\theta},\mathbf{y})}\right|_{\mathbf{x}_{-i}=\mathbf{x}_i^*(x_i,\boldsymbol{\theta})}$$

where $\tilde{\pi}_G(\mathbf{x}_{-i}|x_i,\boldsymbol{\theta},\mathbf{y})$ is the Laplace Gaussian approximation to $\pi(\mathbf{x}_{-i}|x_i,\boldsymbol{\theta},\mathbf{y})$ and $\mathbf{x}_i^*(x_i,\boldsymbol{\theta})$ is its modal configuration. Then the integrals are calculated numerically, to obtain the approximated posterior distributions $\tilde{\pi}(x_i|\boldsymbol{\theta},\mathbf{y})$ and $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$.

Lastly, we apply the main results to the observations of a vector field describing the velocity of the wind, which is crucial in the weather forecast. This application illustrates the importance of our approach, because instead of dealing directly with the solutions of fluid dynamic equations, we model the complexity of the phenomena by fitting a statistical spatial model to the real dataset, using Bayesian Inference.

2

Background

This chapter presents the basic mathematical and statistical concepts and results, needed for the work that is developed in the following chapters. In Section 2.1, we present the functional spaces to be used throughout the work. Namely, we recall the notion of Sobolev spaces and state some of their properties. We introduce the Schwartz space $\mathcal{S}$, the Schwartz distribution space $\mathcal{S}^*$, and the definition of $*$-weak convergence in $\mathcal{S}^*$. Then we recall the definition of generalized random function, and the definition of the white noise as a generalized random function. In Section 2.2 we introduce the principles of Bayesian inference, how Bayes' Theorem is used to calculate the posterior distributions $\pi(\theta|y)$, and hierarchical models. Finally, in Section 2.3, we present an overview of what is Spatial Statistics and the main subjects it addresses, specifying the kind of problem that we are interested in this work.

2.1 Functional Spaces

Consider a measurable space $(\Omega, \mathcal{A})$, where Ω is a set, and $\mathcal{A}$ is a $\sigma - algebra$. We define *measurable function* in the following way.

Definition 2.1.1. *Given a measurable space* $(\Omega, \mathcal{A})$, *and* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *where* $\mathcal{B}$ *is the Borel* σ-*algebra, we say that a function* $f : \Omega \to \mathbb{R}$ *is* measurable *if*

$$f^{-1}(B) := \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{A}, \quad \textit{for any} \quad B \in \mathcal{B}(\mathbb{R}).$$

Consider now a measure space $(\Omega, \mathcal{A}, \mu)$, where μ is a measure defined in the σ-algebra $\mathcal{A}$,

$$\mu : \mathcal{A} \to [0, +\infty].$$

We assume that the measure space is complete. The sets $A \in \mathcal{A}$ such that $\mu(A) = 0$ are called *null sets*. We say that a certain property holds almost everywhere (a.e.) if it holds for every $\omega \in \Omega$, except for a null measure set.

Definition 2.1.2. *A measurable function* $f : \Omega \to \mathbb{R}$ *is said to be* integrable *if the integral of its absolute value is finite,*

$$\|f\|_1 = \int_\Omega |f|\, d\mu < +\infty .$$

In that case, we say that $f \in L^1(\Omega)$.

We will use the notation $L^1(\Omega)$, for the set of all integrable functions. Let us now define the L^p spaces.

Definition 2.1.3. *Consider* $1 < p < +\infty$*. We define*

$$L^p(\Omega) := \left\{ f : \Omega \to \mathbb{R} \quad | \quad f \quad \textit{is measurable and} \quad |f|^p \in L^1(\Omega) \right\}$$

and

$$\|f\|_{L^p} = \|f\|_p := \left(\int_\Omega |f|^p \, d\mu \right)^{\frac{1}{p}} .$$

Definition 2.1.4. *Considering* $p = +\infty$*, we define*

$$L^\infty(\Omega) := \{ f : \Omega \to \mathbb{R} \quad | \quad f \textit{ is measurable and} \quad \exists C \quad |f(x)| < C \quad \textit{a.e. on } \Omega \}$$

and

$$\|f\|_{L^\infty} = \|f\|_\infty := \inf\{ C : |f(\omega)| < C \quad \textit{a.e. on } \Omega \}.$$

L^p spaces provided with the norms $\|\cdot\|_p$ are Banach spaces, for all $1 \le p \le \infty$ (see [1]). We define in $L^2(\Omega)$ the inner product

$$(f, g) := \int_\Omega f\, g\, d\mu \tag{2.1.1}$$

for all $f, g \in L^2(\Omega)$. $L^2(\Omega)$ provided with the inner product $(\cdot,\cdot)$ defined in (2.1.1) is an Hilbert space.

Now we present a motivation for the definition of the *weak derivative*. For that, consider $\mathbb{R}^d$ provided with the Lebesgue measure, and let $\mathcal{O} \subset \mathbb{R}^d$ be an open set. Let $C_0^\infty(\mathcal{O})$ denote the space of infinitely differentiable functions $\phi : \mathcal{O} \to \mathbb{R}$, with compact support in $\mathcal{O}$. We call *test functions* to a function $\phi \in C_0^\infty(\mathcal{O})$.

Given $f \in C^1(\mathcal{O})$ and $\phi \in C_0^\infty(\mathcal{O})$, from the integration by parts formula, and since ϕ has compact support, we have

$$\int_\mathcal{O} f \frac{\partial \phi}{\partial x_i}\, dx = -\int_\mathcal{O} \frac{\partial f}{\partial x_i} \phi\, dx, \quad \text{for} \quad i = 1, \dots, d. \tag{2.1.2}$$

Furthermore, if $f \in C^k(\mathcal{O})$, and $\alpha = (\alpha_1, \dots, \alpha_d)$ is a multiindex with $|\alpha| = \alpha_1 + \cdots + \alpha_d = k$, applying formula (2.1.2) k times, we obtain

$$\int_\mathcal{O} f D^\alpha \phi\, dx = (-1)^{|\alpha|} \int_\mathcal{O} D^\alpha f \phi\, dx \tag{2.1.3}$$

where

$$D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}.$$

Let $L^1_{loc}(\mathbb{O})$ be the set of locally integrable functions f, that is,

$$\int_K |f| dx < +\infty$$

for any compact set $K \subset \mathbb{O}$. Notice that the left side of equation (2.1.3) is well defined for any $f \in L^1_{loc}(\mathbb{O})$, so it motivates the following definition of weak derivative.

Definition 2.1.5. *Consider $f, g \in L^1_{loc}(\mathbb{O})$, and α a multiindex. We say that g is the α^{th}-weak derivative of f if*

$$\int_{\mathbb{O}} f D^\alpha \phi \, dx = (-1)^{|\alpha|} \int_{\mathbb{O}} g \phi \, dx.$$

for all test function $\phi \in C_0^\infty(\mathbb{O})$. In that case, we write $g = D^\alpha f$.

Proposition 2.1.6. *The weak α^{th}-partial derivative of f, if it exists, is uniquely defined, up to a zero measure set.*

For more information on weak derivatives, see [20].

Now we introduce Sobolev spaces. This spaces have an important role in the study of partial differential equations, and were introduced by S.L. Sobolev in the 1930's. In this framework partial differential equations are interpreted as operators defined in some function space, and the derivatives are interpreted in the weak sense. Generally speaking, as the elements of a Sobolev space admit weak derivatives of several orders, that are L^p functions, those function spaces turn out to be specially suited to solve partial differential equation problems, in the weak sense. It is important to mention that, when a weak solution is regular enough, it coincides with the strong solution. Roughly speaking, a strong solution has enough regularity for the derivatives to exist in the classical sense.

Definition 2.1.7. *Consider $1 \leq p \leq +\infty$. Let k be a nonnegative integer, and let $\mathbb{O} \subseteq \mathbb{R}^d$ be an open set. The Sobolev space $W^{k,p}(\mathbb{O})$ is the set of all $f \in L^p(\mathbb{O})$ such that, for each multiindex α with $|\alpha| \leq k$, the weak derivative $D^\alpha f$ exists and belongs to $L^p(\mathbb{O})$.*

The norm $\|\cdot\|_{k,p}$ on $W^{k,p}(\mathbb{O})$ is defined in the following way.

Definition 2.1.8. *Consider $f \in W^{k,p}(\mathbb{O})$, we define the norm as*

$$\|f\|_{W^{k,p}} = \|f\|_{k,p} := \begin{cases} \left(\sum_{|\alpha| \leq k} \int_{\mathbb{O}} |D^\alpha f|^p \, dx \right)^{1/p} & \text{if} \quad 1 \leq p < +\infty \\ \sum_{|\alpha| \leq k} \|D^\alpha f\|_\infty & \text{if} \quad p = +\infty \end{cases} \tag{2.1.4}$$

Theorem 2.1.9. *Consider $1 \leq p \leq +\infty$, and let k be a nonnegative integer. Then $W^{k,p}$ endowed with the norm $\|\cdot\|_{k,p}$ is a Banach space.*

For more information on Sobolev spaces, we refer [1], [7], [20].

If $p = 2$, we write

$$H^k(\mathcal{O}) = W^{k,2}(\mathcal{O}), \quad \text{for} \quad k = 0, 1, \dots .$$

Notice that $H^0(\mathcal{O}) = L^2(\mathcal{O})$. Moreover, we have that $H^k(\mathcal{O})$ is an Hilbert space (see [1]).

Following [23], we present an embedding theorem for Sobolev spaces, needed in Chapter 3.

Let $(\mathcal{U}, \|\cdot\|_{\mathcal{U}})$ and $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ be two normed spaces. We say that $\mathcal{V}$ is continuously embedded in $\mathcal{U}$ if

(i) $\mathcal{V}$ is a vector subspace of $\mathcal{U}$, and

(ii) The identity operator $I : \mathcal{V} \to \mathcal{U}$, defined by $Ix = x$ is continuous.

According to [1], since I is linear, (ii) is equivalent to the following,

$$\|x\|_{\mathcal{U}} \leq C\|x\|_{\mathcal{V}}$$

for all $x \in \mathcal{V}$, and some constant C. We say that $\mathcal{V}$ is compactly embedded in $\mathcal{U}$ if the embedding operator is compact.

We say that a domain $\mathcal{O} \subseteq \mathbb{R}^d$ has locally Lipschitz boundary Γ if, for every $x \in \Gamma$, there exists a neighbourhood V_x such that $\Gamma \cap V_x$ is the graph of a Lipschitz continuous function.

Theorem 2.1.10. *Let $\mathcal{O} \subseteq \mathbb{R}^d$ be a bounded domain with locally Lipschitz boundary Γ. Let $p \in \mathbb{R}$ with $1 \leq p < \infty$, and let $m, n \in \mathbb{N}$ with $n \leq m$. Then the following embedding is compact*

$$W^{m,p}(\mathcal{O}) \hookrightarrow W^{n,q}(\mathcal{O})$$

for all $q \in \mathbb{R}$ such that

$$\begin{cases} 1 \leq q < dp/(d-(m-n)p)\,, & \text{if} \quad d > (m-n)p \\ 1 \leq q < \infty\,, & \text{if} \quad d = (m-n)p \end{cases}$$

Recall the Schwartz space $\mathcal{S} = C_0^\infty(\mathcal{O})$, of all infinitely differentiable functions with compact support, which is also known as the space of *Schwartz test functions*. Next we introduce the concept of *generalized function*, also known as *distribution*, which is a linear continuous functional defined on the Schwartz test functions space, $\mathcal{S}$.

We say that a sequence of test functions $(f_n)_{n\in\mathbb{N}}$, converges to f in $\mathcal{S}$, if there exists a compact $K \subset \mathbb{R}^d$ such that the support of f_n is contained in K,

$$\operatorname{supp}(f_n) := \overline{\{x \in \mathbb{R}^d : f_n \neq 0\}} \subset K, \quad \text{for all } n \in \mathbb{N}\,,$$

and all derivatives

$$\partial^k f_n \to \partial^k f$$

converge uniformly, where ∂^k is the multi index derivative,

$$\partial^k = \frac{\partial^{|k|}}{\partial^{k_1} x_1 \dots \partial^{k_d} x_d}, \quad k = (k_1, \dots, k_d), \quad \text{and} \quad |k| = k_1 + \dots + k_d\,.$$

This convergence is equivalent to the convergence of all derivatives with respect to the norm in $L^2(\mathbb{R}^d)$ (see [34]),

$$\|\partial^k f_n - \partial^k f\|_2^2 := \int_{\mathbb{R}^d} |\partial^k f_n - \partial^k f|^2 \, dt \to 0 \, .$$

Moreover, the Hilbert space $L^2(\mathbb{R}^d)$ with the inner product

$$(f,g) = \int_{\mathbb{R}^d} f(x) g(x) \, dx$$

is the closure of $\mathcal{S}$ with respect to the L^2-norm, $L^2(\mathbb{R}^p) = [\mathcal{S}]$, (see [34]).

We say that a linear functional

$$T : \mathcal{S} \to \mathbb{R}$$

is continuous if

$$T(f_n) \to T(f)$$

for all $(f_n)_{n\in\mathbb{N}} \subset \mathcal{S}$ such that $f_n \to f$ in $\mathcal{S}$. We consider the Borel topology in $\mathbb{R}$, $\mathcal{B}(\mathbb{R})$.

Definition 2.1.11. *A* generalized function *is a linear continuous functional $T : \mathcal{S} \to \mathbb{R}$, also called* distribution. *We write the value of T evaluated at $f \in \mathcal{S}$ as $T(f) = (T, f)$.*

The space of all generalized functions defined on $\mathcal{S}$ is also known as the *Schwartz distribution space*, or the *dual space* of $\mathcal{S}$, and is denoted by $\mathcal{S}^*$. An example of a general function in $\mathcal{S}$ is the Dirac delta function, which is defined by

$$(\delta, f) = f(0) \, ,$$

for every $f \in \mathcal{S}$. Formally, we can write the Dirac delta function as

$$\delta(x) = \begin{cases} +\infty & \text{if} \quad x = 0 \\ 0 & \text{if} \quad x \neq 0 \end{cases} .$$

Also, we can formally write

$$(\delta, f) = \int_{\mathbb{R}} \delta(x) f(x) \, dx \, .$$

For a given x_0, we have the following

$$\delta(x - x_0) = \begin{cases} +\infty & \text{if} \quad x = x_0 \\ 0 & \text{if} \quad x \neq x_0 \end{cases} .$$

We also denote

$$\delta_{x_0}(x) = \delta(x - x_0)$$

so δ_{x_0} is the distribution given by

$$(\delta_{x_0}, f) = f(x_0)$$

and formally we can write

$$(\delta_{x_0}, f) = \int_{\mathbb{R}} \delta_{x_0}(x) f(x)\, dx = \int_{\mathbb{R}} \delta(x - x_0) f(x)\, dx\,.$$

The dual space $\mathcal{S}^*$ can be endowed with a topology induced by $\mathcal{S}$, called the *$*$-weak topology*. The convergence in $\mathcal{S}^*$ with respect to the $*$-weak topology is defined in the following way.

Definition 2.1.12. *Given a sequence $(T_n)_{n\in\mathbb{N}} \subset \mathcal{S}^*$, we say that T_n converges $*$-weakly to $T \in \mathcal{S}^*$ if*

$$(T_n, f) \to (T, f)$$

for all $f \in \mathcal{S}$. In that case, we write $T_n \rightharpoonup T$ $*$-weakly.

Let us consider a normed vector space $(\mathcal{U}, \|\cdot\|_{\mathcal{U}})$. We call the *dual space* of $\mathcal{U}$ to the set of all linear continuous functionals $T : \mathcal{U} \to \mathbb{R}$, with respect to the topology induced by $\|\cdot\|_{\mathcal{U}}$. The dual space is denoted by $\mathcal{U}^*$, and induces a topology in $\mathcal{U}$, called the *weak topology*, defined as follows.

Definition 2.1.13. *We say that a sequence $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$* converges weakly *to $u \in \mathcal{U}$ if*

$$T(u_n) \to T(u) \quad \textit{for all} \quad T \in \mathcal{U}^*.$$

In that case, we write $u_n \rightharpoonup u$.

Next, we recall the concept of *generalized random function*. This concept is important, because our partial differential equations are stochastic. Instead of a deterministic force, we consider a white noise process, and following [34], the white noise is defined as a generalized random function. We also need to recall the *convolution* of two functions in $L^2(\mathbb{R}^d)$, and *convolution* of a generalized function with an element of $L^2(\mathbb{R}^d)$.

Since $L^2(\mathbb{R}^d) = [\mathcal{S}]$, given $f \in L^2(\mathbb{R}^d)$, there exists a sequence $(f_n)_{n\in\mathbb{N}} \in \mathcal{S}$ such that $f = \lim f_n$ with respect to the L^2-norm. Consider $T \in \mathcal{S}^*$. T is a linear continuous functional, so (T, f_n) converges in $\mathbb{R}$. Therefore, the equality

$$(T, f) := \lim (T, f_n)$$

defines the value of T for any $f \in L^2(\mathbb{R}^d)$. Notice that this definition does not depend on the sequence $(f_n)_{n\in\mathbb{N}}$. Next, we present the definition of *generalized random function*.

Definition 2.1.14. *Consider a probability space $(\Omega, \mathcal{A}, P)$, and a vector normed vector space $(\mathcal{U}, \|\cdot\|_{\mathcal{U}})$. A* generalized random function *ξ is is a random variable with values in $\mathcal{U}^*$,*

$$\xi : \Omega \to \mathcal{U}^*,$$

then we have that $(\xi, f) : \Omega \to \mathbb{R}$ is a random variable.

Recall that, given $f,g \in L^2(\mathbb{R}^d)$, the convolution $f*g$ is defined in the following way,

$$f*g(s) := \int_{\mathbb{R}^d} f(s-s')g(s')ds' = \int_{\mathbb{R}^d} f(s')g(s-s')ds' \,.$$

Moreover, for a generalized random function ξ, and $g \in L^2(\mathbb{R}^d)$, the convolution $\xi * g$ is defined as

$$(\xi * g, f) := (\xi, f * g_\flat) \tag{2.1.5}$$

for all $f \in L^2(\mathbb{R}^d)$, where $g_\flat(s) = g(-s)$. Following [34], we present the definition of *white noise* which corresponds to a generalized random function.

Definition 2.1.15. *The* white noise *is a generalized random function $\mathcal{W}$ such that*

$$\mathbb{E}(\mathcal{W}, f) = 0 \quad \textit{and} \quad \mathbb{E}[(\mathcal{W}, f,)(\mathcal{W}, g)] = (\delta, f * g_\flat)$$

for all $f,g \in L^2(\mathbb{R}^d)$.

Notice the following,

$$\mathbb{E}[(\mathcal{W}, f)(\mathcal{W}, g)] = (\delta, f*g_\flat) = f*g_\flat(0) = \int_{\mathbb{R}^d} f(s')g_\flat(-s')ds' = \int_{\mathbb{R}^d} f(s')g(s')ds' = (f,g) \,.$$

Consequently,

$$\mathbb{E}[(\mathcal{W}, f)^2] = \|f\|_2^2$$

for all $f \in L^2(\mathbb{R}^d)$. Moreover, by (2.1.5),

$$\begin{aligned}
\mathbb{E}[(\mathcal{W}, f)(\mathcal{W}, g)] &= (\delta, f*g_\flat) \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \delta(y-x)g(y)dy \right) f(x)dx \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \delta(y-x)f(x)g(y)\,dydx
\end{aligned}$$

where $f,g \in L^2(\mathbb{R}^d)$.

2.2 Bayesian Statistics

2.2.1 Bayes' Theorem

According to [31], it is believed that Bernoulli (1713) was one of the first authors to define probability as degree of credibility. De Morgan (1847) states that probability identifies with some degree of credibility, and degrees of probability can be measured. Ramsey (1926) believes that if an individual believes the odds of some proposition are $r:s$, then the degree of credibility of such proposition is $r/(r+s)$. The idea of probability as degree of credibility plays an important role in Bayesian inference.

Consider a probability space $(\Omega, \mathscr{A}, P)$ where Ω is a non empty set called the *sample space*, $\mathscr{A}$ is σ-algebra of measurable subsets of Ω, and P is a probability measure. Consider a partition $\{A_i : i \in \mathcal{I}\}$ of Ω, where $\mathcal{I}$ is a countable set of indexes. Consider an event $B \in \mathcal{A}$, such that $P(B) > 0$. Bayesian Inference considers that the events A_i are *hypothesis* or *states of nature* to which we attribute *a priori* degrees of credibility, in other words, *a priori* probabilities $P(A_i)$, for $i \in \mathcal{I}$. These prior probabilities are result of previous knowledge about the phenomena that is being studied. After knowing that event B occurred, the probabilities of A_i, for $i \in \mathcal{I}$, are updated. They are provided with new *a posteriori* probabilities via Bayes' Theorem,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i\in\mathcal{I}} P(B|A_i)P(A_i)} \,.$$

Consider that we observe a random variable Y, with a density function $\pi(y|\theta)$, where $\theta \in \Theta$ is an unknown parameter or vector of unknown parameters. Bayesian inference states that all unknown quantities must be quantified in terms of probabilities, therefore, θ is considered a random variable, with prior distribution $\pi(\theta)$ that must be specified, and incorporates the prior knowledge about the experience. The purpose is to obtain the posterior distribution of the parameters, $\pi(\theta|y)$, via Bayes's Theorem. For simplicity, we will use the expression density function for continuous and discrete variables. From conditional probability,

$$\pi(y,\theta) = \pi(y|\theta)\pi(\theta) = \pi(\theta|y)\pi(y) \,,$$

so Bayes' Theorem for density functions can be stated as

$$\pi(\theta|y) = \frac{\pi(y|\theta)\pi(\theta)}{\pi(y)} = \frac{\pi(y|\theta)\pi(\theta)}{\int_\Theta \pi(y|\theta)\pi(\theta)\,d\theta} \,, \quad \theta \in \Theta \,.$$

2.2.2 Hierarchical models

Hierarchical models are designed to combine different sources of information, at different levels. For example, the observations are assumed to depend on some set of parameters, which are treated as random variables themselves, depending on some other set of hyper-parameters. In this work, we are interested in the following hierarchical model. Consider the problem of modelling $\mathbf{y} = (y_1, \ldots, y_n)$, where each observation y_i follows some distribution $\pi(y_i|\mu_i, \boldsymbol{\theta}_2)$, depending on μ_i, and also a vector of parameters $\boldsymbol{\theta}_2$. Moreover, the parameter μ_i is considered to have a prior distribution $\pi(\mu_i|\boldsymbol{\theta}_1)$, that depends on a vector of hyperparameters $\boldsymbol{\theta}_1$, and finally, for simplicity, we call vector of all parameters to $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, which is considered to follow a prior distribution $\pi(\boldsymbol{\theta})$. The prior distributions reflect the prior knowledge about the parameters of the phenomena that is being modelled. The model described can be formulated as

$$\begin{aligned} y_i|\mu_i, \boldsymbol{\theta} &\sim \pi(y_i|\mu_i, \boldsymbol{\theta}), \quad i \in \{1, \ldots, n\} \\ \mu_i|\boldsymbol{\theta} &\sim \pi(\mu_i|\boldsymbol{\theta}) \\ \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}) \,. \end{aligned} \tag{2.2.1}$$

In the context of Bayesian Inference, parameters are also random variables, so to standardize the notation, capital letters are left out. Let $\mu_i = E[y_i]$, and assume that $g(\mu_i) = \eta_i$ for some given *link function* $g(\cdot)$ and an *additive predictor* η_i. The additive predictor is defined as

$$\eta_i = \alpha + \sum_{l=1}^{L} z_i^l \beta_l + \sum_{k=1}^{K} f_k(w_i^k), \quad i \in \{1, \ldots, n\}. \tag{2.2.2}$$

The scalar α stands for the intercept, coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_L)$ account for the linear effects of covariates $\mathbf{z} = (\mathbf{z}^1, \ldots, \mathbf{z}^L)$, and $\mathbf{f} = (\mathbf{f_1}(\cdot), \ldots, \mathbf{f_K}(\cdot))$, are unknown functions of covariates $\mathbf{w} = (\mathbf{w}^1, \ldots, \mathbf{w}^K)$, that may represent, for instance, non-linear effects or spatial random effects. The intercept α, the coefficients $\boldsymbol{\beta}$, and the random effects $\mathbf{f}$, are called *latent variables* of the model, since they are not directly observed. The vector of all latent variables is

$$\mathbf{x} = (\alpha, \boldsymbol{\beta}, \mathbf{f}), \tag{2.2.3}$$

that is,

$$\mathbf{x} = (\alpha, \beta_1, \beta_2, \ldots, \beta_L, f_1(w_1^1), ..., f_1(w_n^1), f_2(w_1^2), ..., f_2(w_n^2), \ldots, f_K(w_1^K), ..., f_K(w_n^K)) . \tag{2.2.4}$$

2.3 Spatial Statistics

Spatial Statistics is a subject of Statistics that developed mainly due to the need of study data collected at different locations, presenting spatial dependency. The advance of computational tools and technology, allowed researchers to obtain and process spatial and spatio-temporal data, and to develop spatial models in a wide range of subjects: climatology, social science, epidemiology, and others (see [5] and [16]).

Spatial Statistics is composed by three branches, namely *Geostatistics*, *areal data*, and *point processes* (see [16]).

Geostatistics is characterized by the study of a certain phenomena, observed at a finite number of locations, $\{y(s_1), \ldots, y(s_m)\}$, with $s_i \in D \subset \mathbb{R}^d$, for $i = 1, \ldots, m$. It allows the locations s_i to vary continuously through the domain D. An example could be precipitation measurements at a finite number of locations, in limited domain D.

Areal data deals with observations in subregions of a partition of the domain $D \subset \mathbb{R}^d$, $\{y(A_1), \ldots, y(A_m)\}$, where $A_i \subset D \subset \mathbb{R}^d$, for $i = 1, \ldots, m$, with $A_i \cap A_j = \emptyset$, for all $i \neq j$. We could think of $y(A_i)$ as the observed number of patients with some disease in each region $A_i \subset D$. Usually we choose a point within A_i to represent A_i. These points will form a lattice, that can be regular or irregular, that is why in some literature, areal data is also called *lattice data*.

Finally, point processes deal with data that are given by random locations of occurrence of some phenomena. For instance, we could think of the locations of pine trees in a certain forest. The questions that arise in that case could be: is there clustering of pine trees? Or is it totally random? Or can we detect any regularity or pattern in their locations? Also, we may attribute to each configuration of locations $\{s_1, \cdots, s_m\}$ an observation of a characteristic of the phenomena that we are observing, for instance, the diameter of the pine tree in location s_i, represented by a vector $\{z(s_1), \ldots, z(s_m)\}$. In that case, we call the point process *marked point process*.

In this work we are interested in Geostatistics phenomena.

3

Third grade non-Newtonian fluid

In this chapter we study the deterministic third grade fluid equation, perturbed by a multiplicative white noise $\frac{d\mathcal{W}_t}{dt}$,

$$d(v(y)) = \Big(-\nabla p + \nu \Delta y - (Y \cdot \nabla) v - \sum_j v^j \nabla y^j + (\alpha_1 + \alpha_2) \operatorname{div}\big(A^2\big) + \beta \operatorname{div}\big(|A|^2 A\big) + U\Big)\, dt + \sigma(t,y)\, d\mathcal{W}_t \,, \tag{3.0.1}$$

where y is the velocity field of the fluid, U is a body force and

$$A(y) = \nabla y + \nabla y^{\top}, \qquad v(y) = y - \alpha_1 \Delta y \,. \tag{3.0.2}$$

According to [19] and [21],

$$\nu \geq 0, \qquad \alpha_1 \geq 0, \qquad \beta \geq 0, \qquad |\alpha_1 + \alpha_2| \leq \sqrt{24\nu\beta} \tag{3.0.3}$$

in order to obtain compatibility between the motion of the fluid and thermodynamic laws. We prove the existence and uniqueness of the solution for the stochastic third grade fluid equation, extending the deterministic results obtained in [8] and [10]. We follow the methods developed in [6], which have been successfully applied to the stochastic second grade fluids in [12], and [33]. Considering an appropriate Galerkin basis, we construct a sequence of approximate solutions, then we deduce uniform estimates in order to get weak convergence of a subsequence. The weak limit is projected on the finite n−dimensional space, and then we show that the difference between the projection of the the weak limit, and the finite dimensional Galerkin approximations, converges strongly to zero up to a certain stopping time. The results of this chapter have been published in the article [14].

3.1 Functional setting and notations

We consider the stochastic third grade fluid equation (3.0.1) in a bounded, not axisymmetric and simply connected domain $\mathcal{O}$ of $\mathbb{R}^2$ with a sufficiently regular boundary Γ, and supplemented with a Navier slip boundary condition, which reads

$$\begin{cases} d(v(Y)) = \Big(-\nabla p + \nu\Delta Y - (Y\cdot\nabla)v - \sum_j v^j \nabla Y^j + (\alpha_1+\alpha_2)\operatorname{div}\Big(A^2\Big) \\ \qquad\qquad +\beta \operatorname{div}\Big(|A|^2 A\Big) + U\Big)dt + \sigma(t,Y)\,d\mathcal{W}_t, & \text{in } \mathcal{O}\times(0,T), \\ \operatorname{div} Y = 0 & \text{in } \mathcal{O}\times(0,T), \\ Y\cdot \mathrm{n} = 0, \qquad (\mathrm{n}\cdot D(Y))\cdot\tau = 0 & \text{on } \Gamma\times(0,T), \\ Y(0) = Y_0 & \text{in } \mathcal{O}, \end{cases} \tag{3.1.1}$$

where $Y = (Y_, .Y_2)$ is the velocity field of the fluid, ∇Y is its Jacobian matrix, $D(Y) = \frac{\nabla Y + (\nabla Y)^\top}{2}$, $A = A(Y) = 2D(Y)$, $v(Y) = Y - \alpha_1 \Delta Y$ and the constants ν, α_1, α_2, β verify (3.0.3). The stochastic perturbation is defined by

$$\sigma(t,Y)\,d\mathcal{W}_t = \sum_{k=1}^{m} \sigma^k(t,Y)\,d\mathcal{W}_t^k,$$

where the diffusion coefficient

$$\sigma(t,Y) = (\sigma^1(t,Y),\dots,\sigma^m(t,Y))$$

satisfies certain growth assumptions defined below, and $\mathcal{W}_t = (\mathcal{W}_t^1,\dots,\mathcal{W}_t^m)$ is a standard $\mathbb{R}^m$-valued Wiener process defined on a complete probability space $(\Omega,\mathcal{A},P)$ endowed with a filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ such that $\mathcal{W}_t$, $t\in[0,T]$, is adapted to $\mathcal{F}$. We assume that $\mathcal{F}_0$ contains every P-null subset of Ω.

Let us introduce the Helmholtz projector $\mathbb{P} : L^2(\mathcal{O}) \longrightarrow H$, which is the linear bounded operator characterized by the following L^2–orthogonal decomposition

$$v = \mathbb{P}v + \nabla\phi, \qquad \phi \in H^1(\mathcal{O}).$$

We consider the boundary conditions to define the following Hilbert spaces,

$$\begin{aligned} H &= \Big\{y \in L^2(\mathcal{O}) \mid \operatorname{div} y = 0 \text{ in } \mathcal{O} \text{ and } y\cdot\mathrm{n} = 0 \text{ on } \Gamma\Big\}, \\ V &= \Big\{y \in H^1(\mathcal{O}) \mid \operatorname{div} y = 0 \text{ in } \mathcal{O} \text{ and } y\cdot\mathrm{n} = 0 \text{ on } \Gamma\Big\}, \\ W &= \Big\{y \in V\cap H^2(\mathcal{O}) \mid (\mathrm{n}\cdot D(y))\cdot\tau = 0 \text{ on } \Gamma\Big\}. \end{aligned} \tag{3.1.2}$$

On H we consider the L^2–inner product $(\cdot,\cdot)$ and the associated norm $\|\cdot\|_2$, and we define the following inner products

$$(u,z)_V := (v(u),z) = (u,z) + 2\alpha_1\,(Du,Dz), \tag{3.1.3}$$

$$(u,z)_W := (u,z)_V + (\mathbb{P}v(u),\mathbb{P}v(z)), \tag{3.1.4}$$

We denote by $\|\cdot\|_V$ and $\|\cdot\|_W$ the norms induced by these inner product, respectively.

Notice that V is a subspace of $H^1(\mathcal{O})$, so it is also endowed with the norm $\|\cdot\|_{H^1}$, nevertheless both norms $\|\cdot\|_{H^1}$ and $\|\cdot\|_V$ on the space V are equivalent. Similarly $W \subset H^2(\mathcal{O})$ and the norms $\|\cdot\|_W$ and $\|\cdot\|_{H^2}$ are equivalent on W.

Let us define the trilinear functional

$$b(\phi, z, y) = (\phi \cdot \nabla z, y), \qquad \forall \phi, z, y \in V. \tag{3.1.5}$$

Once ϕ is divergence free and $(\phi \cdot n) = 0$ on Γ, applying integration by parts we obtain

$$b(\phi, z, y) = -b(\phi, y, z). \tag{3.1.6}$$

We need the following inequalities. The Korn inequality states the following,

$$\|y\|_{W^{1,p}} \leq K_1(p)\big(\|y\|_p + \|A(y)\|_p\big), \quad \forall y \in V, \quad p \geq 2, \tag{3.1.7}$$

while the Poincaré inequality establishes the following,

$$\|y\|_2 \leq C\|\nabla y\|_2, \quad \forall y \in V. \tag{3.1.8}$$

Moreover, for non axisymmetric bounded domains, we have the following version of the Korn inequality (see [18])

$$\|\nabla y\|_2 \leq K_2(\mathcal{O})\|A(y)\|_2, \quad \forall y \in V. \tag{3.1.9}$$

The Sobolev embedding $H^1(\mathcal{O}) \hookrightarrow L^4(\mathcal{O})$ and (3.1.8) give

$$\|y\|_4 \leq K_3\|\nabla y\|_2, \quad \forall y \in V.$$

Combining this inequality with (3.1.9), we get

$$\|y\|_4 \leq K_3\|\nabla y\|_2 \leq K_3K_2(\mathcal{O})\|A(y)\|_2. \tag{3.1.10}$$

Due to the embedding $L^4(\mathcal{O}) \hookrightarrow L^2(\mathcal{O})$, we have

$$\|y\|_2 \leq C_*\|y\|_4. \tag{3.1.11}$$

Then (3.1.10), (3.1.11) and (3.1.7) yield the following lemma:

Lemma 3.1.1. *There exists a positive constant K_* such that*

$$\|y\|_{W^{1,4}} \leq K_*\|A(y)\|_4, \quad \forall y \in V. \tag{3.1.12}$$

Also recall the Young's inequality

$$uz \leq \frac{1}{r}u^r + \frac{1}{s}z^s, \quad \forall u, z \geq 0, \quad s, r > 0 \quad \text{such that} \quad \frac{1}{r} + \frac{1}{s} = 1. \tag{3.1.13}$$

Accordingly, for real numbers γ, a, b, x such that $0 \leq \gamma < a$ and $b, x \geq 0$, we have the algebraic relation

$$\forall \delta > 0, \qquad bx^\gamma \leq C(\gamma, a, b, \delta) + \delta x^a, \tag{3.1.14}$$

that will be used widely in this chapter.

Let us mention that through this chapter, C will represent a generic constant, whose value can change from line to line. To explicitly write its dependence with respect of some parameters $\alpha_1, \ldots, \alpha_k$, we also write $C(\alpha_1, \ldots, \alpha_k)$ instead of C.

3.2 Preliminary results

We consider the following auxiliary modified Stokes problem with Navier boundary condition

$$\begin{cases} \tilde{f} - \alpha_1 \Delta \tilde{f} = f - \nabla p, & \operatorname{div} \tilde{f} = 0 \quad \text{in } \mathcal{O}, \\ \tilde{f} \cdot \mathrm{n} = 0, & (\mathrm{n} \cdot D(\tilde{f})) \cdot \tau = 0 \quad \text{on } \Gamma. \end{cases} \tag{3.2.1}$$

We recall from [11] that assuming $f \in H^m(\mathcal{O})$, $m = 0, 1$, the problem (3.2.1) has a solution $(\tilde{f}, p) \in H^{m+2}(\mathcal{O}) \times H^{m+1}(\mathcal{O})$ verifying

$$\|\tilde{f}\|_{H^2} \leq C\|f\|_2. \tag{3.2.2}$$

According to the definition of the inner product (3.1.3), we have

$$(\tilde{f}, h)_V = (f, h), \quad \forall h \in V. \tag{3.2.3}$$

In the next two lemmas, we establish properties of the nonlinear terms that will be useful in Section 3.3 to identify the weak limits of the nonlinear terms of the equation. Let us introduce the operators

$$S(y) := \beta \left(|A(y)|^2 A(y) \right), \tag{3.2.4}$$

$$N(y) := \alpha_1 \left(y \cdot \nabla A(y) + (\nabla y)^\top A(y) + A(y) \nabla y \right) - \alpha_2 (A(y))^2. \tag{3.2.5}$$

Here we collect important inequalities from [8] related with the nonlinear terms.

Lemma 3.2.1. *For any $\epsilon > 0$ and $y \in W$, we have*

$$\left| (\alpha_1 + \alpha_2) \int_{\mathcal{O}} \operatorname{div}(A^2) \cdot y \right| \leq \epsilon \|A^2\|_2^2 + \frac{(\alpha_1 + \alpha_2)^2}{16\epsilon} \|A\|_2^2, \tag{3.2.6}$$

where $A = A(y)$.

Proof. The integration by parts gives

$$(\alpha_1 + \alpha_2) \int_{\mathcal{O}} \operatorname{div}(A^2) \cdot y = (\alpha_1 + \alpha_2) \int_{\Gamma} \left(\mathrm{n} \cdot A^2 \right) \cdot y - (\alpha_1 + \alpha_2) \int_{\mathcal{O}} A^2 \cdot \nabla y. \tag{3.2.7}$$

Due to the boundary conditions $y = (y \cdot \tau)\tau$ and $(\mathrm{n} \cdot A) \cdot \tau = 0$ on Γ, we obtain

$$\begin{aligned} (\mathrm{n} \cdot A^2) \cdot y &= (y \cdot \tau)(\mathrm{n} \cdot A^2) \cdot \tau = (y \cdot \tau)((\mathrm{n} \cdot A) \cdot A) \cdot \tau \\ &= (y \cdot \tau)\left[((\mathrm{n} \cdot A) \cdot \mathrm{n})((\mathrm{n} \cdot A) \cdot \tau) + ((\mathrm{n} \cdot A) \cdot \tau)((\tau \cdot A) \cdot \tau)\right] = 0. \end{aligned}$$

Using the symmetry of A, we derive

$$(\alpha_1 + \alpha_2) \int_{\mathcal{O}} \operatorname{div}(A^2) \cdot y = -\frac{1}{2} (\alpha_1 + \alpha_2) \int_{\mathcal{O}} A^2 \cdot A. \tag{3.2.8}$$

Therefore, the Hölder and the Young inequalities give (3.2.6).

■

Considering a small change in estimate (35) of [8], we collect the following estimates.

Lemma 3.2.2 (see [8], relations (33)-(36))**.** *For each $y \in W$ and any $\epsilon, \delta > 0$, the following estimates are valid*

$$\Big(\mathrm{div}\Big(|A|^2A\Big), \mathbb{P}v(y)\Big) \leq -\frac{1}{2}\|A\|_4^4 - \frac{\alpha_1}{2}\||A||\nabla A|\|_2^2 - \frac{\alpha_1}{4}\|\nabla(|A|^2)\|_2^2 + 3\epsilon\||A||\nabla^2 y|\|_2^2 + 5\epsilon\|A\|_{12}^4 + 3\epsilon\|y\|_{H^1}^4 + C(\epsilon)\|y\|_{H^1}^2\|y\|_{H^2}^2, \quad (3.2.9)$$

$$(\alpha_1+\alpha_2)\Big(\mathrm{div}\Big(A^2\Big), \mathbb{P}v(y)\Big) \leq \epsilon\||A||\nabla^2 y|\|_2^2 + C(\epsilon)\|y\|_W^2, \quad (3.2.10)$$

$$-\Big((y\cdot\nabla)v + \sum_j v^j\nabla y^j, \mathbb{P}v(y)\Big) \leq 4\epsilon\||A||\nabla^2 y|\|_2^2 + C(\epsilon,\delta)\|y\|_W^2 + C(\epsilon)\|y\|_\infty\|y\|_W^2 + \delta\|y\|_{W^{1,4}}^4, \quad (3.2.11)$$

where $A = A(y)$ and $v = v(y)$.

Lemma 3.2.3. *For any $y, \hat{y} \in W$, we have*

$$\langle \mathrm{div}(S(\hat{y}) - S(y)), \hat{y} - y\rangle = -\frac{\beta}{4}\int_{\mathcal{O}}(|\hat{A}|^2 - |A|^2)^2 - \frac{\beta}{4}\int_{\mathcal{O}}(|\hat{A}|^2 + |A|^2)|A(\hat{y}-y)|^2, \quad (3.2.12)$$

where $A = A(y)$ and $\hat{A} = A(\hat{y})$.

Proof. Integrating by parts, we write

$$\langle \mathrm{div}(S(\hat{y}) - S(y)), \hat{y} - y\rangle = \beta\int_\Gamma \Big(\mathrm{n}\cdot(|\hat{A}|^2\hat{A} - |A|^2A)\Big)\cdot(\hat{y}-y) - \frac{\beta}{2}\int_{\mathcal{O}}(|\hat{A}|^2\hat{A} - |A|^2A)\cdot A(\hat{y}-y) = I_{11} + I_{12}. \quad (3.2.13)$$

Using the boundary conditions, we deduce that

$$I_{11} = \beta\int_\Gamma((\hat{y}-y)\cdot\tau)\Big[|\hat{A}|^2(\mathrm{n}\cdot\hat{A})\cdot\tau - |A|^2(\mathrm{n}\cdot A)\cdot\tau\Big] = 0. \quad (3.2.14)$$

Standard algebraic computations yield

$$\begin{aligned} I_{12} &= -\frac{\beta}{2}\int_{\mathcal{O}}(|\hat{A}|^2\hat{A} - |A|^2A)\cdot A(\hat{y}-y) \\ &= -\frac{\beta}{2}\int_{\mathcal{O}}(|\hat{A}|^2\hat{A} - |A|^2A)\cdot A(\hat{y}-y) \\ &= -\frac{\beta}{4}\int_{\mathcal{O}}(|\hat{A}|^2 - |A|^2)^2 - \frac{\beta}{4}\int_{\mathcal{O}}(|\hat{A}|^2 + |A|^2)|A(\hat{y}-y)|^2. \end{aligned} \quad (3.2.15)$$

■

Lemma 3.2.4. *For any y, $\hat{y} \in W$, the following estimate holds*

$$\begin{aligned} (\mathrm{div}\,(N(\hat{y}) - N(y)), \hat{y}-y) &= (N(\hat{y}) - N(y), \nabla(\hat{y}-y)) \\ &\leq 3\epsilon\int_{\mathcal{O}}|A(\hat{y}-y)|^2\Big(|A|^2 + |\hat{A}|^2\Big) + \frac{C}{\epsilon}\int_{\mathcal{O}}|\nabla(\hat{y}-y)|^2 \\ &\quad + \frac{C}{1-\lambda}\epsilon^{\frac{\lambda-1}{\lambda+3}}\|\hat{y}-y\|_{H^1}^{\frac{4(\lambda+1)}{\lambda+3}}\|y\|_{H^2}^{\frac{4}{\lambda+3}} \quad \text{for any} \quad \epsilon > 0,\ \lambda \in]0,1[, \end{aligned} \quad (3.2.16)$$

where $A = A(y)$ and $\hat{A} = A(\hat{y})$.

Proof. The divergence theorem gives

$$(\operatorname{div}(N(\hat{y})-N(y)),\hat{y}-y)=(N(\hat{y})-N(y),\nabla(\hat{y}-y))-\int_{\Gamma}[(N(\hat{y})-N(y))\mathrm{n}]\cdot(\hat{y}-y). \quad (3.2.17)$$

The relation (3.2.16) is proved in [10] for the case $\mathcal{O}=\mathbb{R}^2$ (domain without boundary), in [8] it is verified that the boundary term vanishes.

■

Lemma 3.2.5. *For any y, $\hat{y}$, $\phi \in W$, we have*

$$\left|\langle \operatorname{div}(S(y)-S(\hat{y}),\phi\rangle\right| \leq C\|y\|_W^2\|y-\hat{y}\|_V\|\phi\|_W + C\|\hat{y}\|_W\left\||A|^2-|\hat{A}|^2\right\|_2\|\phi\|_W, \quad (3.2.18)$$

where $A=A(y)$ and $\hat{A}=A(\hat{y})$.

Proof. Using the Hölder inequality, and the Sobolev injections $H^1(\mathcal{O})) \hookrightarrow L^p(\mathcal{O})$ for $p<\infty$ and $H^2(\mathcal{O})\hookrightarrow L^\infty(\mathcal{O})$, we derive

$$\begin{aligned}
\left|\langle \operatorname{div}(S(y)-S(\hat{y})),\phi\rangle\right| &= \left|\beta\int_{\mathcal{O}}(|A|^2A-|\hat{A}|^2\hat{A})\cdot\nabla\phi\right| \\
&= \left|\beta\int_{\mathcal{O}}\Big(|A|^2(A-\hat{A})+\hat{A}(|A|^2-|\hat{A}|^2)\Big)\cdot\nabla\phi\right| \\
&\leq C\||A|^2\|_4\|A(y-\hat{y})\|_2\|\nabla\phi\|_4 + C\|\hat{A}\|_4\left\||A|^2-|\hat{A}|^2\right\|_2\|\nabla\phi\|_4 \\
&\leq C\|y\|_{H^2}^2\|y-\hat{y}\|_{H^1}\|\phi\|_{H^2} + C\|\hat{y}\|_{H^2}\left\||A|^2-|\hat{A}|^2\right\|_2\|\phi\|_{H^2}. \quad (3.2.19)
\end{aligned}$$

■

Lemma 3.2.6. *For any y, $\hat{y}$, $\phi\in W$, the following inequality holds*

$$\begin{aligned}
\left|\langle \operatorname{div}(N(\hat{y})-N(y)),\phi\rangle\right| &\leq C\epsilon\left\|A(y-\hat{y})\sqrt{|A|^2+|\hat{A}|^2}\right\|_2\|\phi\|_V \\
&\quad + C\|\hat{y}-y\|_V\,(\|y\|_W+\|\hat{y}\|_W)\,\|\phi\|_W. \quad (3.2.20)
\end{aligned}$$

where $A=A(y)$ and $\hat{A}=A(\hat{y})$.

Proof. Here we apply the same reasoning that is done in [10] to show the property (3.2.16).

$$\begin{aligned}
\langle \operatorname{div}(N(\hat{y})-N(y)),\phi\rangle &= \langle N(\hat{y})-N(y),\nabla\phi\rangle = \frac{1}{2}\langle N(\hat{y})-N(y),A(\phi)\rangle \\
&= \frac{\alpha_2}{2}\int_{\mathcal{O}}(A^2-\hat{A}^2)\cdot A(\phi) - \frac{\alpha_1}{2}\int_{\mathcal{O}}(y\cdot\nabla A-\hat{y}\cdot\nabla\hat{A})\cdot A(\phi) \\
&\quad -\frac{\alpha_1}{2}\int_{\mathcal{O}}\Big((\nabla y)^\top A + A\nabla y-(\nabla\hat{y})^\top\hat{A}-\hat{A}\nabla\hat{y}\Big)\cdot A(\phi) = I_1+I_2+I_3. \quad (3.2.21)
\end{aligned}$$

$$|I_1| \leq C \left\| |A(y-\hat{y})| \sqrt{|A|^2 + |\hat{A}|^2} \right\|_2 \|A(\phi)\|_2. \tag{3.2.22}$$

Next, we use the properties of the trilinear form, as well as the Hölder inequality, and the Sobolev injections $H^1(\mathcal{O}) \hookrightarrow L^4(\mathcal{O})$ and $H^2(\mathcal{O}) \hookrightarrow L^\infty(\mathcal{O})$ in order to deduce that

$$\begin{aligned}
|I_2| &\leq C \left| b(y, A, A(\phi)) - b(\hat{y}, \hat{A}, A(\phi)) \right| \\
&\leq C \left| b(y, A - \hat{A}, A(\phi)) \right| + \left| b(y - \hat{y}, \hat{A}, A(\phi)) \right| \\
&= C \left| b(y, A(\phi), A - \hat{A}) \right| + \left| b(y - \hat{y}, \hat{A}, A(\phi)) \right| \\
&\leq C \|y\|_\infty \|A(\phi)\|_{H^1} \|A - \hat{A}\|_2 + \|y - \hat{y}\|_4 \|\hat{A}\|_{H^1} \|A(\phi)\|_4 \\
&\leq C \|y - \hat{y}\|_{H^1} \left(\|y\|_{H^2} + \|\hat{y}\|_{H^2} \right) \|\phi\|_{H^2}.
\end{aligned} \tag{3.2.23}$$

For I_3 we have the same estimate as for I_1, namely

$$|I_3| \leq C \left\| |A(y-\hat{y})| \sqrt{|A|^2 + |\hat{A}|^2} \right\|_2 \|A(\phi)\|_2. \tag{3.2.24}$$

■

3.3 Existence of strong solution

This section establishes the main results of the article. More precisely the solution of the equation is constructed through the finite dimensional Galerkin approximation method. We first deduce key uniform estimates for the finite dimensional approximations in order to get a weakly convergent sequence. Next, with the help of a suitable stopping time, and using the structure of the equation, we improve the convergence results. Finally, with these new convergence results, we will be able to identify the nonlinear terms of the equation.

Let us to introduce the notion of the solution.

Definition 3.3.1. *Let $U \in L^2(\Omega \times (0,T), L^2(\mathcal{O}))$ and $Y_0 \in L^2(\Omega, W)$. Then a stochastic process $Y \in L^2(\Omega, L^\infty(0,T;W))$ is a strong solution of (3.1.1), if the following equation holds*

$$\begin{aligned}
(v(Y(t)), \phi) &= \int_0^t \Big[-2\nu \left(D(Y), D(\phi) \right) + ((Y \cdot \nabla)\phi, v(Y)) - \sum_j \left(v^j(Y) \nabla Y^j, \phi \right) \Big] ds \\
&\quad - \int_0^t \left((\alpha_1 + \alpha_2) \left(A^2 \right) + \beta \left(|A|^2 A \right), \nabla \phi \right) ds \\
&\quad + (v(Y(0)), \phi) + \int_0^t (U(s), \phi)\, ds + \int_0^t (\sigma(s, Y(s)), \phi)\, d\mathcal{W}_s
\end{aligned} \tag{3.3.1}$$

for a.e. $(\omega, t) \in \Omega \times [0,T]$ and for all $\phi \in V$, where the stochastic integral is defined by

$$\int_0^t (\sigma(s, Y(s)), \phi)\, d\mathcal{W}_s = \sum_{k=1}^m \int_0^t \left(\sigma^k(s, Y(s)), \phi \right) d\mathcal{W}_s^k.$$

Now we state the main result. Assume that the diffusion coefficient $\sigma = \left(\sigma^1, \dots, \sigma^m\right)$: $[0,T] \times V \to (L^2(\mathcal{O}))^m$ is Lipschitz in the second variable and verifies a growth condition, i.e., there exist positive constants L, K and $0 \le \gamma < 2$ such that

$$\left\|\sigma(t,y)\right\|_2^2 \le L(1 + \|y\|_{W^{1,4}}^{\gamma}), \quad \forall y \in W^{1,4}(\mathcal{O}) \cap V, \tag{3.3.2}$$

$$\left\|\sigma(t,y) - \sigma(t,z)\right\|_2^2 \le K \left\|y - z\right\|_V^2, \qquad \forall y,z \in V,\ t \in [0,T], \tag{3.3.3}$$

where

$$\left\|\sigma(t,y)\right\|_2^2 := \sum_{i=1}^{m} \left\|\sigma^i(t,y)\right\|_2^2.$$

In addition, we define

$$|(\sigma(t,y),v)| := \left(\sum_{k=1}^{m} \left(\sigma^k(t,y),v\right)^2\right)^{1/2}, \quad \forall v \in L^2(\mathcal{O}).$$

Moreover, we take $p \ge 6$ and suppose that the initial condition Y_0 and the force U satisfy the following,

$$Y_0 \in L^p(\Omega, W), \quad \text{and there exists } \lambda > 0 \text{ such that} \quad \mathbb{E}e^{\lambda\left(\int_0^T \|U\|_2^2 ds + \|Y_0\|_V^2\right)} < \infty. \tag{3.3.4}$$

Theorem 3.3.2. *Assume* (3.3.2)-(3.3.4). *Then there exists a unique solution* Y *to equation* (3.1.1) *which belongs to*

$$L^p(\Omega, L^\infty(0,T;W)).$$

In order to show the existence of the solution, we apply the *Galerkin's approximation method* for an appropriate basis. We recall that the injection operator $I : W \hookrightarrow V$ being a compact operator guarantees the existence of a basis $\{e_i\} \subset W$ of eigenfunctions to the problem

$$(v, e_i)_W = \lambda_i (v, e_i)_V, \qquad \forall v \in W, \quad i \in \mathbb{N}, \tag{3.3.5}$$

which is an orthonormal basis in V and an orthogonal basis in W. In addition the sequence $\{\lambda_i\}$ of the corresponding eigenvalues fulfils the properties: $\lambda_i > 0$, $\forall i \in \mathbb{N}$, and $\lambda_i \to \infty$ as $i \to \infty$. Since the ellipticity of the equation (3.3.5) increases the regularity of their solutions (see [9]), we may consider $\{e_i\} \subset H^4$.

We consider the finite dimensional space $W_n = \operatorname{span}\{e_1, \dots, e_n\}$, and introduce the Faedo-Galerkin approximation of the system (3.1.1). Namely, we look for a solution to the following stochastic differential equation

$$\begin{cases} d\,(v_n, \phi) = \Big(\nu \Delta Y_n - (Y_n \cdot \nabla) v_n - \sum_j v_n^j \nabla Y_n^j + (\alpha_1 + \alpha_2)\operatorname{div}\left(A_n^2\right) \\ \qquad\qquad + \beta \operatorname{div}\left(|A_n|^2 A_n\right) + U, \phi\Big) dt + (\sigma(t,Y_n), \phi)\, d\mathcal{W}_t, \qquad \forall \phi \in W_n, \\ Y_n(0) = Y_{n,0}, \end{cases} \tag{3.3.6}$$

where

$$Y_n(t) = \sum_{j=1}^{n} c_j^n(t) e_j.$$

Here $Y_{n,0}$ denotes the projection of the initial condition Y_0 onto the space W_n, $v_n = Y_n - \alpha_1 \Delta Y_n$ and $A_n = \nabla Y_n + (\nabla Y_n)^\top$.

Due to the relation (3.3.5), the sequence $\{\tilde{e}_j = \frac{1}{\sqrt{\lambda_j}} e_j\}$ is an orthonormal basis for W and

$$Y_{n,0} = \sum_{j=1}^{n} \left(Y_0, e_j\right)_V e_j = \sum_{j=1}^{n} \left(Y_0, \tilde{e}_j\right)_W \tilde{e}_j,$$

The Parseval's identity yields

$$\|Y_n(0)\|_V \leq \|Y_0\|_V \quad \text{and} \quad \|Y_n(0)\|_W \leq \|Y_0\|_W.$$

The equation (3.3.6) can be written as a system of stochastic ordinary differential equations in $\mathbb{R}^n$ with locally Lipschitz nonlinearities. From classical results there exists a local-in-time solution Y_n that is an adapted stochastic process with values in $C([0,T_n], W_n)$.

The existence of a global-in-time solution follows from the uniform estimates on $n = 1, 2, \ldots$, that will be deduced in the next lemma (a similar reasoning can be found in [2], [11], [33]).

Lemma 3.3.3. *Let us assume* (3.3.2)-(3.3.4). *Then the problem (3.3.6) admits a unique solution $Y_n \in L^2(\Omega, L^\infty(0,T;W))$, which verifies the following estimates*

$$\begin{aligned} \mathbb{E} \sup_{s\in[0,t]} \|Y_n(s)\|_V^2 \quad &+ \quad 8\nu \mathbb{E} \int_0^t \|DY_n\|_2^2 \, ds + \frac{\beta}{4} \mathbb{E} \int_0^t \|A_n\|_4^4 \, ds \\ &\leq \quad C\left(1 + \mathbb{E}\|Y_0\|_V^2 + \mathbb{E}\|U\|^2_{L^2(0,t;L^2(\mathcal{O}))}\right), \quad \forall t \in [0,T], \end{aligned} \tag{3.3.7}$$

$$\mathbb{E}\|Y_n\|^4_{L^4(0,t;W^{1,4}(\mathcal{O}))} \leq C\left(1 + \mathbb{E}\|Y_0\|_V^2 + \mathbb{E}\|U\|^2_{L^2(0,t;L^2(\mathcal{O}))}\right), \quad \forall t \in [0,T], \tag{3.3.8}$$

where C is a positive constant independent of n.

Proof. For each $n \in \mathbb{N}$, we define the following sequence of stopping times

$$\tau_M^n = \inf\{t \geq 0 : \|Y_n(t)\|_V \geq M\} \wedge T_n, \quad M \in \mathbb{N}.$$

Let us set

$$f(Y_n) := \nu \Delta Y_n - (Y_n \cdot \nabla) v_n - \sum_j v_n^j \nabla Y_n^j + (\alpha_1 + \alpha_2) \operatorname{div}\left(A_n^2\right) + \beta \operatorname{div}\left(|A_n|^2 A_n\right) + U. \tag{3.3.9}$$

Using (3.1.3), and considering in (3.3.6) the test functions $\phi = e_i$, $i = 1, \ldots, n$, we write

$$d\,(Y_n, e_i)_V = (f(Y_n), e_i)\, dt + (\sigma(t, Y_n), e_i)\, d\mathcal{W}_t. \tag{3.3.10}$$

Applying the Itô formula, we deduce

$$d\,(Y_n,e_i)_V^2 = 2\,(Y_n,e_i)_V\,(f(Y_n),e_i)\,dt + 2\,(Y_n,e_i)_V\,(\sigma(t,Y_n),e_i)\,d\mathcal{W}_t + |(\sigma(t,Y_n),e_i)|^2\,dt.$$

Summing over $i=1,\ldots,n$, we derive

$$d\,\|Y_n\|_V^2 = 2\,(f(Y_n),Y_n)\,dt + 2\,(\sigma(t,Y_n),Y_n)\,d\mathcal{W}_t + \sum_{i=1}^{n}|(\sigma(t,Y_n),e_i)|^2\,dt. \tag{3.3.11}$$

We have

$$\begin{aligned}(f(Y_n),Y_n) &= -2\nu\,\|DY_n\|_2^2 - \Big((Y_n\cdot\nabla)v_n + \sum_j v_n^j\nabla Y_n^j, Y_n\Big)\\ &\quad + \Big((\alpha_1+\alpha_2)\operatorname{div}\big(A_n^2\big),Y_n\Big) + \Big(\beta\operatorname{div}\big(|A_n|^2A_n\big),Y_n\Big) + (U,Y_n)\\ &= I_1+I_2+I_3+I_4+I_5.\end{aligned} \tag{3.3.12}$$

By the symmetry of the trilinear functional (3.1.5), we obtain

$$\begin{aligned}I_2 &= -\Big((Y_n\cdot\nabla)v_n + \sum_j v_n^j\nabla Y_n^j, Y_n\Big) = -b(Y_n,v_n,Y_n) - b(Y_n,Y_n,v_n)\\ &= -b(Y_n,v_n,Y_n) + b(Y_n,v_n,Y_n) = 0.\end{aligned} \tag{3.3.13}$$

Taking into account the boundary conditions $Y_n=(Y_n\cdot\tau)\tau$, $(n\cdot A_n)\cdot\tau=0$ on Γ and the symmetry of ∇Y_n, the divergence theorem gives

$$\begin{aligned}I_4 &= \int_{\mathcal{O}}\beta\operatorname{div}\big(|A_n|^2A_n\big)\cdot Y_n = \beta\int_\Gamma |A_n|^2(Y_n\cdot\tau)(n\cdot A_n)\cdot\tau - \beta\int_{\mathcal{O}}|A_n|^2A_n\cdot\nabla Y_n\\ &= -\frac{\beta}{2}\|A_n\|_4^4.\end{aligned} \tag{3.3.14}$$

Taking $\epsilon=\frac{\beta}{4}$ in (3.2.6), we obtain

$$|I_3| \le \frac{\beta}{4}\|A_n^2\|_2^2 + \frac{(\alpha_1+\alpha_2)^2}{4\beta}\|A_n\|_2^2. \tag{3.3.15}$$

In addition, we have

$$|I_5| = |(U,Y_n)| \le \frac{1}{2}\|Y_n\|_2^2 + \frac{1}{2}\|U\|_2^2. \tag{3.3.16}$$

Therefore, introducing (3.3.13)-(3.3.16) in (3.3.11), we derive

$$\begin{aligned}d\,\|Y_n\|_V^2 + \frac{\beta}{2}\|A_n\|_4^4\,dt + 4\nu\,\|DY_n\|_2^2\,dt &\le \frac{(\alpha_1+\alpha_2)^2}{2\beta}\|A_n\|_2^2\,dt\\ &\quad + (\|U\|_2^2+\|Y_n\|_2^2)\,dt + 2\,(\sigma(t,Y_n),Y_n)\,d\mathcal{W}_t + \sum_{i=1}^{n}|(\sigma(t,Y_n),e_i)|^2\,dt.\end{aligned} \tag{3.3.17}$$

We write

$$\begin{aligned}d\,\|Y_n\|_V^2 + \frac{\beta}{2}\|A_n\|_4^4\,dt + 4\nu\,\|DY_n\|_2^2\,dt &\le C(\beta,\alpha_1,\alpha_2)\|Y_n\|_V^2\,dt\\ &\quad + \|U\|_2^2\,dt + 2\,(\sigma(t,Y_n),Y_n)\,d\mathcal{W}_t + \sum_{i=1}^{n}|(\sigma(t,Y_n),e_i)|^2\,dt.\end{aligned} \tag{3.3.18}$$

Denoting by $\tilde{\sigma}_n$ the solution of the generalized Stokes problem (3.2.1) for $f = \sigma(t, Y_n)$, we have

$$(\tilde{\sigma}_n, e_i)_V = (\sigma(t, Y_n), e_i) \quad \text{for } i = 1, \dots, n,$$

then (3.3.2), (3.1.12) and Young's inequality give

$$\begin{aligned}\sum_{i=1}^{n} |(\sigma(t, Y_n), e_i)|^2 &= \|\tilde{\sigma}_n\|_V^2 \leq C\|\sigma(t, Y_n)\|_2^2 \leq CL(1 + \|Y_n\|_{W^{1,4}}^{\gamma}) \\ &\leq CL + CL(K_*)^{\gamma}\|A_n\|_4^{\gamma} \leq C(L, \gamma, \beta, K_*) + \frac{\beta}{4}\|A_n\|_4^4. \end{aligned} \tag{3.3.19}$$

For any $t \in [0, T]$, integrating the inequality (3.3.18) on $(0, s)$, $s \in [0, \tau_M^n \wedge t]$ and using (3.3.19), we derive

$$\begin{aligned}\|Y_n(s)\|_V^2 + \frac{\beta}{4}\int_0^s \|A_n\|_4^4\, dr + 4\nu \int_0^s \|DY_n\|_2^2\, dr &\leq \|Y_n(0)\|_V^2 + C(L, \gamma, \beta, K_*, T) \\ + C(\beta, \alpha_1, \alpha_2)\int_0^s \|Y_n\|_V^2\, dr + \int_0^s \|U\|_2^2\, dr &+ 2\int_0^s (\sigma(r, Y_n), Y_n)\, d\mathcal{W}_r. \end{aligned} \tag{3.3.20}$$

On the other hand, the Burkholder-Davis-Gundy inequality, (3.3.2), (3.1.12) and the Young inequality yield

$$\begin{aligned}&\mathbb{E} \sup_{s\in[0,\tau_M^n\wedge t]} \left| \int_0^s (\sigma(r, Y_n), Y_n)\, dW_r \right| \leq C\mathbb{E}\left(\int_0^{\tau_M^n \wedge t} |(\sigma(s, Y_n), Y_n)|^2\, ds \right)^{\frac{1}{2}} \\ &\leq C\mathbb{E}\left(\int_0^{\tau_M^n \wedge t} \|\sigma(s, Y_n)\|_2^2 \|Y_n\|_2^2\, ds \right)^{\frac{1}{2}} \leq C\mathbb{E}\left(\int_0^{\tau_M^n \wedge t} L(1 + \|Y_n\|_{W^{1,4}}^{\gamma})\|Y_n\|_V^2\, ds \right)^{\frac{1}{2}} \\ &\leq \frac{1}{4}\mathbb{E} \sup_{s\in[0,\tau_M^n\wedge t]} \|Y_n\|_V^2 + C^2LT + \mathbb{E}\int_0^{\tau_M^n\wedge t} C^2L(K_*)^{\gamma}\|A_n\|_4^{\gamma}\, ds \\ &\leq \frac{1}{4}\mathbb{E} \sup_{s\in[0,\tau_M^n\wedge t]} \|Y_n\|_V^2 + C(L, \gamma, \beta, K_*, T) + \frac{\beta}{16}\mathbb{E}\int_0^{\tau_M^n\wedge t} \|A_n\|_4^4\, ds. \end{aligned} \tag{3.3.21}$$

Taking the supremum on $s \in [0, \tau_M^n \wedge t]$ and the expectation in (3.3.20) and incorporating the estimate (3.3.21), we obtain

$$\begin{aligned}&\frac{1}{2}\mathbb{E} \sup_{s\in[0,\tau_M^n\wedge t]} \|Y_n(s)\|_V^2 + 4\nu\mathbb{E}\int_0^{\tau_M^n\wedge t} \|DY_n\|_2^2\, ds + \frac{\beta}{8}\mathbb{E}\int_0^{\tau_M^n\wedge t} \|A_n\|_4^4\, dr \\ &\leq C(L, \gamma, \beta, K_*, T) + \mathbb{E}\|Y_0\|_V^2 + \mathbb{E}\int_0^t \|U\|_2^2\, ds + C(\beta, \alpha_1, \alpha_2)\mathbb{E}\int_0^t \sup_{r\in[0,\tau_M^n\wedge s]} \|Y_n(r)\|_V^2\, ds. \end{aligned} \tag{3.3.22}$$

Then the function

$$f(t) = \frac{1}{2}\mathbb{E} \sup_{s\in[0,\tau_M^n\wedge t]} \|Y_n(s)\|_V^2 + 4\nu\mathbb{E}\int_0^{\tau_M^n\wedge t} \|DY_n\|_2^2\, ds + \frac{\beta}{8}\mathbb{E}\int_0^{\tau_M^n\wedge t} \|A_n\|_4^4\, ds$$

fulfils the Gronwall's inequality

$$f(t) \leq C + \mathbb{E}\|Y_0\|_V^2 + \mathbb{E}\int_0^t \|U\|_2^2\, ds + C\int_0^t f(s) ds,$$

which implies

$$\mathbb{E} \sup_{s\in[0,\tau_M^n\wedge t]} \|Y_n(s)\|_V^2 + 8\nu\mathbb{E}\int_0^{\tau_M^n\wedge t} \|DY_n\|_2^2\,ds + \frac{\beta}{4}\mathbb{E}\int_0^{\tau_M^n\wedge t} \|A_n\|_4^4\,ds \leq C\Big(1+\mathbb{E}\|Y_0\|_V^2+\mathbb{E}\|U\|^2_{L^2(0,t;L^2(\mathcal{O}))}\Big). \quad (3.3.23)$$

Then there exists a constant C independent of M and n such that

$$\mathbb{E} \sup_{s\in[0,\tau_M^n\wedge t]} \|Y_n(s)\|_V^2 \leq C, \quad \forall t\in[0,T]. \quad (3.3.24)$$

Let us fix $n\in\mathbb{N}$, writing

$$\mathbb{E} \sup_{s\in[0,\tau_M^n\wedge T]} \|Y_n(s)\|_V^2 = \mathbb{E}\Bigg(\sup_{s\in[0,\tau_M^n\wedge T]} 1_{\{\tau_M^n<T\}}\,\|Y_n(s)\|_V^2\Bigg) + \mathbb{E}\Bigg(\sup_{s\in[0,\tau_M^n\wedge T]} 1_{\{\tau_M^n\geq T\}}\,\|Y_n(s)\|_V^2\Bigg)$$
$$\geq \mathbb{E}\Bigg(\sup_{s\in[0,\tau_M^n]} 1_{\{\tau_M^n<T\}}\,\|Y_n(s)\|_V^2\Bigg) \geq M^2 P\Big(\tau_M^n<T\Big), \quad (3.3.25)$$

we deduce that $P\Big(\tau_M^n<T\Big)\leq \frac{C}{M^2}$. This means that $\tau_M^n\to T$ in probability, as $M\to\infty$. Then there exists a subsequence $\{\tau_{M_k}^n\}$ of $\{\tau_M^n\}$ (that may depend on n) such that

$$\tau_{M_k}^n \to T \quad \text{a.e. as } k\to\infty.$$

Since $\tau_{M_k}^n \leq T_n \leq T$, we deduce that $T_n = T$, so Y_n is a global-in-time solution of the stochastic differential equation (3.3.6). In addition for fixed n, the monotonicity of the sequence $\big\{\tau_M^n\big\}$ allows to apply the monotone convergence theorem in order to pass to the limit, as $M\to\infty$, in the inequality (3.3.23) in order to obtain (3.3.7).

$$\mathbb{E} \sup_{s\in[0,t]} \|Y_n(s)\|_V^2 + 8\nu\mathbb{E}\int_0^{t} \|DY_n\|_2^2\,ds + \frac{\beta}{4}\mathbb{E}\int_0^{t} \|A_n\|_4^4\,ds \leq C\Big(1+\mathbb{E}\|Y_0\|_V^2+\mathbb{E}\|U\|^2_{L^2(0,t;L^2(\mathcal{O}))}\Big). \quad (3.3.26)$$

This inequality gives

$$\mathbb{E}\int_0^t \|A_n\|_4^4\,ds \leq C(\beta,\alpha_1,\alpha_2)\Big(1+\mathbb{E}\|Y_0\|_V^2+\mathbb{E}\|U\|^2_{L^2(0,t;L^2(\mathcal{O}))}\Big), \quad \forall t\in[0,T], \quad (3.3.27)$$

that together with Lemma 3.1.1 yields

$$\mathbb{E}\|Y_n\|^4_{L^4(0,t;W^{1,4}(\mathcal{O}))} = \mathbb{E}\int_0^t \|Y_n\|^4_{W^{1,4}}ds \leq (K_*)^4\mathbb{E}\int_0^t \|A_n\|_4^4\,ds$$
$$\leq C(\beta,\alpha_1,\alpha_2)\Big(1+\mathbb{E}\|Y_0\|_V^2+\mathbb{E}\|U\|^2_{L^2(0,t;L^2(\mathcal{O}))}\Big), \quad \forall t\in[0,T]. \quad (3.3.28)$$

The Hölder's inequality also gives

$$\mathbb{E}\|Y_n\|_{L^4(0,t;W^{1,4}(\mathcal{O}))} \leq C(\beta,\alpha_1,\alpha_2)\Big(1+\mathbb{E}\|Y_0\|_V^2+\mathbb{E}\|U\|^2_{L^2(0,t;L^2(\mathcal{O}))}\Big)^{\frac{1}{4}}, \quad \forall t\in[0,T]. \quad (3.3.29)$$

■

Lemma 3.3.4. *Assume* (3.3.2)-(3.3.4). *Then we have*

$$\mathbb{E}e^{\frac{\lambda\beta}{16(K_*)^4}\int_0^t \|Y_n\|_{W^{1,4}}^4 ds} < C\,\mathbb{E}e^{\lambda\left(\int_0^T \|U\|_2^2 ds + \|Y_0\|_V^2\right)}, \quad \forall t \in [0,T], \tag{3.3.30}$$

where C is a positive constant independent of n, and K_* is defined by (3.1.12).

Proof. Let us consider the inequality (3.3.20) and write

$$\|Y_n(t)\|_V^2 + \frac{\beta}{4}\int_0^t \|A_n\|_4^4 ds + 4\nu\int_0^t \|DY_n\|_2^2\, ds \le \|Y_n(0)\|_V^2 + \int_0^t \|U\|_2^2 ds$$
$$+ C(L,\gamma,\beta,K_*,T) + C(\beta,\alpha_1,\alpha_2)\int_0^t \|Y_n\|_V^2\, ds + 2\int_0^t (\sigma(s,Y_n),Y_n)\, d\mathcal{W}_s. \tag{3.3.31}$$

Multiplying by $\frac{\lambda}{2}$ and knowing that $W^{1,4}(\mathcal{O}) \hookrightarrow H^1(\mathcal{O})$, we deduce

$$\frac{\lambda}{2}\|Y_n(t)\|_V^2 + \frac{\lambda\beta}{8}\int_0^t \|A_n\|_4^4 ds + 2\lambda\nu\int_0^t \|DY_n\|_2^2\, ds \le \frac{\lambda}{2}\left(\|Y_n(0)\|_V^2 + \int_0^t \|U\|_2^2 ds\right)$$
$$+ C(L,\gamma,\beta,K_*,T) + C(\beta,\alpha_1,\alpha_2)\int_0^t \|Y_n\|_{W^{1,4}}^2 ds + \lambda\int_0^t (\sigma(s,Y_n),Y_n)\, d\mathcal{W}_s.$$

The Korn inequality (3.1.12) gives

$$\frac{\lambda\beta}{8(K_*)^4}\|Y_n\|_{W^{1,4}}^4 \le \frac{\lambda\beta}{8}\|A_n\|_4^4;$$

therefore we have

$$\frac{\lambda\beta}{8(K_*)^4}\int_0^t \|Y_n\|_{W^{1,4}}^4 ds \le \frac{\lambda}{2}\left(\|Y_n(0)\|_V^2 + \int_0^t \|U\|_2^2 ds\right) + C(L,\gamma,\beta,K_*,T)$$
$$+ C(\beta,\alpha_1,\alpha_2)\int_0^t \|Y_n\|_{W^{1,4}}^2 ds + \lambda\int_0^t (\sigma(s,Y_n),Y_n)\, d\mathcal{W}_s. \tag{3.3.32}$$

Let us notice that with the help of (3.3.2), the Sobolev embedding $W^{1,4}(\mathcal{O}) \hookrightarrow H$ and the Young's inequality, for any $\delta > 0$, we can verify that

$$\int_0^t \lambda^2 (\sigma(s,Y_n),Y_n)^2\, ds \le \int_0^t \lambda^2 \|\sigma(s,Y_n)\|_2^2 \|Y_n\|_2^2 ds \le \int_0^t \lambda^2 L(1+\|Y_n\|_{W^{1,4}}^{\gamma})\|Y_n\|_2^2 ds$$
$$\le \int_0^t \lambda^2 L\|Y_n\|_{W^{1,4}}^2 ds + \int_0^t \lambda^2 L\|Y_n\|_{W^{1,4}}^{\gamma+2} ds$$
$$\le C(\lambda,L,\delta,T) + \frac{\delta}{2}\int_0^t \|Y_n\|_{W^{1,4}}^4 ds;$$

which implies

$$-\frac{\delta}{2}\int_0^t \|Y_n\|_{W^{1,4}}^4 ds - C(\lambda,L,\delta,T) \le -\int_0^t \lambda^2 (\sigma(s,Y_n),Y_n)^2\, ds.$$

Adding this relation to (3.3.32), we write

$$\frac{\lambda\beta}{8(K_*)^4}\int_0^t \|Y_n\|_{W^{1,4}}^4 ds - \frac{\delta}{2}\int_0^t \|Y_n\|_{W^{1,4}}^4 ds \le \frac{\lambda}{2}\left(\|Y_n(0)\|_V^2 + \int_0^t \|U\|_2^2 ds\right)$$
$$+ C(L,\gamma,\beta,K_*,T) + C(\beta,\alpha_1,\alpha_2)\int_0^t \|Y_n\|_{W^{1,4}}^2 ds$$
$$+ \lambda\int_0^t (\sigma(s,Y_n),Y_n)\, d\mathcal{W}_s - \int_0^t \lambda^2 (\sigma(s,Y_n),Y_n)^2\, ds. \tag{3.3.33}$$

Once again, the Young inequality gives

$$C(\beta,\alpha_1,\alpha_2)\int_0^t \|Y_n\|_{W^{1,4}}^2\,ds \le C(\beta,\alpha_1,\alpha_2,\delta)+\frac{\delta}{2}\int_0^t \|Y_n\|_{W^{1,4}}^4\,ds.$$

Introducing this estimate in (3.3.33) and next taking $\delta=\frac{\lambda\beta}{16(K_*)^4}$, it follows that

$$\frac{\lambda\beta}{16(K_*)^4}\int_0^t \|Y_n\|_{W^{1,4}}^4\,ds \le \frac{\lambda}{2}\left(\|Y_n(0)\|_V^2+\int_0^t \|U\|_2^2\,ds\right)+C(\beta,\alpha_1,\alpha_2,T)$$
$$+\lambda\int_0^t (\sigma(s,Y_n),Y_n)\,d\mathcal{W}_s-\int_0^t \lambda^2\left(\sigma(s,Y_n),Y_n\right)^2\,ds. \qquad (3.3.34)$$

Now, we take the exponential, the expectation and the Hölder inequality in order to deduce that

$$\mathbb{E}e^{\frac{\lambda\beta}{16(K_*)^4}\int_0^t \|Y_n\|_{W^{1,4}}^4\,ds} \le C(\beta,\alpha_1,\alpha_2,T)\left(\mathbb{E}e^{\lambda\left(\|Y_0\|_V^2+\int_0^t\|U\|_2^2\,ds\right)}\right)^{\frac{1}{2}}$$
$$\left(\mathbb{E}e^{\int_0^t(2\lambda\sigma(s,Y_n),Y_n)\,d\mathcal{W}_s-\frac{1}{2}\int_0^t(2\lambda\sigma(s,Y_n),Y_n)^2\,ds}\right)^{\frac{1}{2}}.$$

Since the stochastic process inside the second expectation is a supermartingale its expectation is less or equal to 1, hence we obtain (3.3.30).

■

Lemma 3.3.5. *Assume* (3.3.2)-(3.3.4). *Then the unique solution* Y_n *of the problem (3.3.6) verifies the following uniform estimate*

$$\mathbb{E}\sup_{s\in[0,t]}\|Y_n(s)\|_W^p \le C, \qquad \forall t\in[0,T], \qquad (3.3.35)$$

where C *is a positive constant independent of* n.

Proof. For each $n\in\mathbb{N}$, let us consider the sequence of stopping times defined by

$$\tau_M^n=\inf\{t\ge 0:\|Y_n(t)\|_W\ge M\},\quad M\in\mathbb{N}.$$

We introduce the solutions $\tilde{f}_n$ and $\tilde{\sigma}_n$ of (3.2.1) for $f=f_n:=f(Y_n)$ (as in (3.3.9)) and $f=\sigma_n:=\sigma(t,Y_n)$, respectively. Then

$$(\tilde{f}_n,e_i)_V=(f_n,e_i),\qquad (\tilde{\sigma}_n,e_i)_V=(\sigma_n,e_i). \qquad (3.3.36)$$

Therefore

$$d(Y_n,e_i)_V=(\tilde{f}_n,e_i)_V\,dt+(\tilde{\sigma}_n,e_i)_V\,d\mathcal{W}_t.$$

Multiplying by λ_i and using (3.3.5), we obtain

$$d\,(Y_n,e_i)_W=(\tilde{f}_n,e_i)_W\,dt+(\tilde{\sigma}_n,e_i)_W\,d\mathcal{W}_t.$$

The Itô formula gives

$$d\,(Y_n,e_i)_W^2=2\,(Y_n,e_i)_W\,(\tilde{f}_n,e_i)_W\,dt+2\,(Y_n,e_i)_W\,(\tilde{\sigma}_n,e_i)_W\,d\mathcal{W}_t+|(\tilde{\sigma}_n,e_i)_W|^2\,dt.$$

Now, multiplying by $\frac{1}{\lambda_i}$ and summing over $i = 1,\dots,n$, we derive

$$d\,\|Y_n\|_W^2 = 2(\tilde{f}_n, Y_n)_W\,dt + 2(\tilde{\sigma}_n, Y_n)_W\,d\mathcal{W}_t + \sum_{i=1}^{n}\frac{1}{\lambda_i}|(\tilde{\sigma}_n, e_i)_W|^2\,dt, \tag{3.3.37}$$

which is equivalent to

$$d\,\|Y_n\|_W^2 = 2\,[(f_n, Y_n) + (f_n, \mathbb{P}v(Y_n))]\,dt + \|\tilde{\sigma}_n\|_W^2\,dt + 2\,[(\sigma_n, Y_n) + (\sigma_n, \mathbb{P}v(Y_n))]\,d\mathcal{W}_t. \tag{3.3.38}$$

Let us recall from (3.3.12)-(3.3.16) that

$$2(f_n, Y_n) \leq -4\nu\,\|DY_n\|_2^2 - \frac{\beta}{2}\|A_n\|_4^4 + \frac{(\alpha_1+\alpha_2)^2}{2\beta}\|A_n\|_2^2 + \|Y_n\|_2^2 + \|U\|_2^2. \tag{3.3.39}$$

On the other hand, considering the Sobolev inequality

$$\|y\|_6 \leq C_1\|y\|_{H^1} \quad \forall y \in H^1,$$

and using the estimates (3.2.9)-(3.2.11) as in [8], page 373, for $\epsilon = \min\left\{\frac{1}{20(C_1)^2}, \frac{\alpha_1}{40(C_1)^2}, \frac{\beta\alpha_1}{9(3\beta+5)}\right\}$ we derive

$$\begin{aligned}2(f_n, \mathbb{P}v(Y_n)) \leq &-\frac{\beta}{2}\|A_n\|_4^4 - \frac{\alpha_1\beta}{2}\||A_n||\nabla A_n|\|_2^2 - \frac{\alpha_1\beta}{4}\|\nabla(|A_n|^2)\|_2^2\\ &+ C(\nu,\beta,\alpha_1,\delta)\|Y_n\|_W^2 + C(\beta,\alpha_1)\|Y_n\|_\infty\|Y_n\|_W^2\\ &+ C(\beta,\alpha_1)\|Y_n\|_V^2\|Y_n\|_W^2 + 2\delta\|Y_n\|_{W^{1,4}}^4 + \|U\|_2^2.\end{aligned} \tag{3.3.40}$$

The Sobolev inequalities

$$\|y\|_\infty \leq C_2\|y\|_{W^{1,4}}, \quad \|y\|_V \leq C_3\|y\|_{W^{1,4}}, \quad \forall y \in W^{1,4},$$

(3.1.11) and the Young's inequality allow to verify that

$$C(\beta,\alpha_1)\left(\|Y_n\|_\infty + \|Y_n\|_V^2\right)\|Y_n\|_W^2 \leq C(\beta,\alpha_1,\delta)\|Y_n\|_W^2 + 2\delta\|y\|_{W^{1,4}}^4\|Y_n\|_W^2, \quad \forall \delta > 0.$$

Therefore, we have

$$\begin{aligned}2(f_n, \mathbb{P}v(Y_n)) \leq &-\frac{\beta}{2}\|A_n\|_4^4 - \frac{\alpha_1\beta}{2}\||A_n||\nabla A_n|\|_2^2 - \frac{\alpha_1\beta}{4}\|\nabla(|A_n|^2)\|_2^2\\ &+ C(\nu,\beta,\alpha_1,\delta)\|Y_n\|_W^2 + 2\delta\|Y_n\|_{W^{1,4}}^4\|Y_n\|_W^2 + 2\delta\|Y_n\|_{W^{1,4}}^4 + \|U\|_2^2.\end{aligned} \tag{3.3.41}$$

Now, we choose δ such that $2D_1 := 4\delta \leq \frac{\lambda\beta}{16p(K_*)^4}$ and introduce the function

$$\xi_1(t) = \mathrm{e}^{-2D_1\int_0^t \|Y_n\|_{W^{1,4}}^4 ds}.$$

We apply the Itô formula to determine the differential of the product $\xi_1(t)\|Y_n(t)\|_W^2$, namely from the equation (3.3.38) we derive

$$\begin{aligned}d\left(\xi_1(t)\|Y_n\|_W^2\right) = &\,\xi_1(t)\,[2(f_n, Y_n) + 2(f_n, \mathbb{P}v(Y_n))]\,dt + \xi_1(t)\|\tilde{\sigma}_n\|_W^2\,dt\\ &+ \xi_1(t)\,[2(\sigma_n, Y_n) + 2(\sigma_n, \mathbb{P}v(Y_n))]\,d\mathcal{W}_t\\ &- 2D_1\xi_1(t)\,\|Y_n\|_W^2\,\|Y_n\|_{W^{1,4}}^4\,dt.\end{aligned} \tag{3.3.42}$$

Using the Itô formula once again for the function $\theta(x) = x^p$, and integrating on $[0,s]$, $s \le t \wedge \tau_M^n$, $t \in [0,T]$, we deduce

$$\begin{aligned}
\left(\xi_1(s)\|Y_n\|_W^2\right)^p &= \|Y_n(0)\|_W^{2p} + p\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1} \xi_1(r)\,[2(f_n,Y_n) + 2(f_n,\mathbb{P}v(Y_n))]\,dr \\
&+ p\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1} \xi_1(r)\|\tilde{\sigma}_n\|_W^2\,dr \\
&+ p\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1} \xi_1(r)\,[2(\sigma_n,Y_n) + 2(\sigma_n,\mathbb{P}v(Y_n))]\,d\mathcal{W}_r \\
&- 2D_1 p\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1} \xi_1(r)\,\|Y_n\|_W^2\,\|Y_n\|_{W^{1,4}}^4\,dr \\
&+ 2p(p-1)\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-2} (\xi_1(r))^2\,[(\sigma_n,Y_n) + (\sigma_n,\mathbb{P}v(Y_n))]^2\,dr.
\end{aligned} \tag{3.3.43}$$

Next, using (3.3.39) and (3.3.41) to estimate the right hand side, we obtain

$$\begin{aligned}
\left(\xi_1(s)\|Y_n\|_W^2\right)^p &\le \|Y_n(0)\|_W^{2p} + p\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1} \xi_1(r)\Big[C(\nu,\beta,\alpha_1,\delta)\|Y_n\|_W^2 + 2\|U\|_2^2\Big]dr \\
&+ p\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1} \xi_1(r)\Big[D_1\|Y_n\|_{W^{1,4}}^4\|Y_n\|_W^2 + D_1\|Y_n\|_{W^{1,4}}^4\Big]\,dr \\
&+ p\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1} \xi_1(r)\|\tilde{\sigma}_n\|_W^2\,ds \\
&+ 2p\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1} \xi_1(r)\,[(\sigma_n,Y_n) + (\sigma_n,\mathbb{P}v(Y_n))]\,d\mathcal{W}_r \\
&- 2D_1 p\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1} \xi_1(r)\,\|Y_n\|_W^2\,\|Y_n\|_{W^{1,4}}^4\,dr \\
&+ 2p(p-1)\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-2} (\xi_1(r))^2\,[(\sigma_n,Y_n) + (\sigma_n,\mathbb{P}v(Y_n))]^2\,dr.
\end{aligned} \tag{3.3.44}$$

Since $\|Y_n\|_W^{2p-2} \le 1 + \|Y_n\|_W^{2p}$, we deduce

$$\begin{aligned}
\left(\xi_1(s)\|Y_n\|_W^2\right)^p &\le \|Y_n(0)\|_W^{2p} + pC(\nu,\beta,\alpha_1,\delta)\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^p\,dr \\
&+ pD_1\int_0^s (\xi_1(r))^p\,\|Y_n\|_{W^{1,4}}^4\,dr \\
&+ 2p\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1} \xi_1(r)\|U\|_2^2\,dr \\
&+ p\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1} \xi_1(r)\|\tilde{\sigma}_n\|_W^2\,dr \\
&+ 2p\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1} \xi_1(r)\,[(\sigma_n,Y_n) + (\sigma_n,\mathbb{P}v(Y_n))]\,d\mathcal{W}_s \\
&+ 2p(p-1)\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-2} (\xi_1(r))^2\,[(\sigma_n,Y_n) + (\sigma_n,\mathbb{P}v(Y_n))]^2\,dr.
\end{aligned} \tag{3.3.45}$$

Taking into account that $\|\tilde{\sigma}_n\|_W^2 \le C\|\sigma\|_2^2$, using (3.3.2), the embedding $W \hookrightarrow W^{1,4}(\mathcal{O})$ and the Young's inequality, we infer that

$$
\begin{aligned}
& p\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1}\xi_1(r)\|\tilde{\sigma}_n\|_W^2\,dr \\
& + 2p(p-1)\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-2}(\xi_1(r))^2\,[(\sigma_n, Y_n) + (\sigma_n, \mathbb{P}v(Y_n))]^2\,dr \\
& \le C(p,T) + C(p)\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^p dr.
\end{aligned}
$$

On the other hand, the Young's inequality (3.1.13) with $r = \frac{p}{p-1}$ and $0 \le \xi_1(t) \le 1$ give

$$
2p\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1}\xi_1(r)\|U\|_2^2\,dr \le 2(p-1)\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^p dr + 2\int_0^s \|U\|_2^{2p}dr.
$$

Introducing these estimates in (3.3.45), we deduce

$$
\begin{aligned}
\left(\xi_1(s)\|Y_n(s)\|_W^2\right)^p \le \|Y_0\|_W^{2p} + C\Big(\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^p dr + \int_0^s \|Y_n\|_{W^{1,4}}^4\,dr + \int_0^s \|U\|_2^{2p}dr + 1\Big) \\
+ 2p\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1}\xi_1(r)\,[(\sigma(r,Y_n), \mathbb{P}v(Y_n) + Y_n)\,d\mathcal{W}_r]. \qquad (3.3.46)
\end{aligned}
$$

The Burkholder-Davis-Gundy inequality, (3.3.2) and the Young's inequality yield

$$
\begin{aligned}
& \mathbb{E}\sup_{s\in[0,\tau_M^n\wedge t]}\left|\int_0^s \left(\xi_1(r)\|Y_n\|_W^2\right)^{p-1}\xi_1(r)\left((\sigma(r,Y_n), \mathbb{P}v(Y_n) + Y_n)\,d\mathcal{W}_r\right)\right| \\
& \le C\mathbb{E}\Big(\int_0^{\tau_M^n\wedge t}\Big(\left(\xi_1(s)\|Y_n\|_W^2\right)^{p-1}\xi_1(s)\Big)^2\|\sigma(s,Y_n)\|_2^2\|Y_n\|_W^2\,ds\Big)^{\frac{1}{2}} \\
& \le C\sqrt{L}\mathbb{E}\Big(\int_0^{\tau_M^n\wedge t}(\xi_1(s))^{2p}\|Y_n\|_W^{4p-2} + \int_0^{\tau_M^n\wedge t}(\xi_1(s))^{2p}\|Y_n\|_W^{4p-2+\gamma}\Big)^{\frac{1}{2}} \\
& \le C\sqrt{2LT} + C\sqrt{2L}\mathbb{E}\Big(\int_0^{\tau_M^n\wedge t}\left(\xi_1(s)\|Y_n\|_W^2\right)^{2p}ds\Big)^{\frac{1}{2}} \\
& \le C\sqrt{2LT} + \frac{\eta}{2p}\mathbb{E}\Big(\sup_{s\in[0,\tau_M^n\wedge t]}\left(\xi_1(s)\|Y_n\|_W^2\right)^p\Big) \\
& + C(L,\eta,p)\mathbb{E}\int_0^{\tau_M^n\wedge t}\left(\xi_1(s)\|Y_n\|_W^2\right)^p ds, \qquad (3.3.47)
\end{aligned}
$$

for any $\eta > 0$. Here we take $\eta = \frac{1}{2}$. Considering the supremum on $s \in [0, \tau_M^n \wedge t]$ and the expectation in (3.3.46), with the help of (3.3.47) we derive the following Gronwall's inequality

$$
\begin{aligned}
& \frac{1}{2}\mathbb{E}\sup_{s\in[0,\tau_M^n\wedge t]}\left(\xi_1(s)\|Y_n(s)\|_W^2\right)^p \le \|Y_0\|_W^{2p} + C\Big(\int_0^t \mathbb{E}\sup_{r\in[0,\tau_M^n\wedge s]}\left(\xi_1(r)\|Y_n(r)\|_W^2\right)^p ds \\
& + \mathbb{E}\int_0^{\tau_M^n\wedge t}\|Y_n\|_{W^{1,4}}^4\,dr + \mathbb{E}\int_0^{\tau_M^n\wedge t}\|U\|_2^{2p}\,dr + 1\Big).
\end{aligned}
$$

Therefore, we obtain

$$\mathbb{E}\sup_{s\in[0,\tau_M^n\wedge t]}\Big(\xi_1(s)\|Y_n(s)\|_W^2\Big)^p \leq C\bigg(1+\mathbb{E}\int_0^t\|U\|_2^{2p}\,dr+\mathbb{E}\int_0^t\|Y_n\|_{W^{1,4}}^4 dr\bigg). \tag{3.3.48}$$

The estimates (3.3.8) and (3.3.4) yield

$$\mathbb{E}\sup_{s\in[0,\tau_M^n\wedge t]}\Big(\xi_1(s)\|Y_n(s)\|_W^2\Big)^p \leq C$$

with C independent of n and M. We verify that for n fixed, $\tau_M^n \to T$ in probability, as $M\to\infty$. Then, there exists a subsequence $\{\tau_{M_k}^n\}$ of $\{\tau_M^n\}$ (that may depend on n) such that $\tau_{M_k}^n \to T$ for a. e. $\omega\in\Omega$, as $k\to\infty$. Using the monotone convergence theorem, we pass to the limit in (3.3.48) as $k\to\infty$, deriving the estimate

$$\mathbb{E}\sup_{s\in[0,t]}\Big(\xi_1(s)\|Y_n(s)\|_W^2\Big)^p \leq C.$$

The Hölder inequality gives

$$\begin{aligned}
\mathbb{E}\sup_{s\in[0,t]}\|Y_n(s)\|_W^p &\leq \mathbb{E}\bigg[\bigg(\sup_{s\in[0,t]}(\xi_1(s))^{\frac{p}{2}}\|Y_n(s)\|_W^p\bigg)(\xi_1(t))^{-\frac{p}{2}}\bigg]\\
&\leq \bigg(\mathbb{E}\sup_{s\in[0,t]}\Big(\xi_1(s)\|Y_n(s)\|_W^2\Big)^p\bigg)^{\frac{1}{2}}\Big(\mathbb{E}(\xi_1(t))^{-p}\Big)^{\frac{1}{2}}\\
&\leq \sqrt{C}\bigg(\mathbb{E}e^{2pD_1\int_0^t\|Y_n\|_{W^{1,4}}^4 ds}\bigg)^{\frac{1}{2}}.
\end{aligned}$$

Using Lemma 3.3.4, we deduce (3.3.35).

■

3.3.1 Proof of Theorem 3.3.2.

In order to show the existence of the solution to the system (3.1.1) it is convenient to write the equation $(3.1.1)_1$ in the following form (see [10], page 3)

$$d(v(Y)) = \Big(-\nabla p+\nu\Delta Y-(Y\cdot\nabla)Y+\mathrm{div}N(Y)+\mathrm{div}S(Y)+U\Big)dt+\sigma(t,Y)\,d\mathcal{W}_t, \tag{3.3.49}$$

with the operators S and N defined in (3.2.4)-(3.2.5). The corresponding finite dimensional approximation reads

$$d(v(Y_n)) = (-\nabla p_n+\nu\Delta Y_n-(Y_n\cdot\nabla)Y_n+\mathrm{div}\,N(Y_n)+\mathrm{div}\,S(Y_n)+U)\,dt+\sigma(t,Y_n)d\mathcal{W}_t. \tag{3.3.50}$$

The proof of Theorem 3.3.2 is splitted into five steps.

Step 1. Convergences related with the projection operator. Let $P_n : W\to W_n$ be the orthogonal projection defined by

$$P_n y = \sum_{j=1}^n \tilde{c}_j\tilde{e}_j \quad \text{with} \quad \tilde{c}_j = \big(y,\tilde{e}_j\big)_W, \qquad \forall y\in W,$$

where $\{\tilde{e}_j = \frac{1}{\sqrt{\lambda_j}} e_j\}_{j=1}^{\infty}$ is the orthonormal basis of W. It is easy to check that

$$P_n y = \sum_{j=1}^{n} c_j e_j \quad \text{with} \quad c_j = \big(y, e_j\big)_V, \qquad \forall y \in W.$$

By Parseval's identity we have that

$$\|P_n y\|_V \le \|y\|_V, \quad \forall y \in V,$$

$$\|P_n y\|_W \le \|y\|_W \qquad \text{and} \qquad P_n y \longrightarrow y \qquad \text{strongly in } W, \quad \forall y \in W.$$

Considering an arbitrary $Z \in L^q(\Omega \times (0,T), W)$, we have

$$\|P_n Z\|_W \le \|Z\|_W \qquad \text{and} \qquad P_n Z(\omega, t) \to Z(\omega, t) \qquad \text{strongly in } W,$$

which are valid for P-a.e. $\omega \in \Omega$ and a.e. $t \in (0,T)$. Hence Lebesgue's dominated convergence theorem and the inequality

$$\|Z\|_V \le C\|Z\|_W \qquad \text{for any } Z \in W$$

imply

$$\begin{aligned} P_n Z &\longrightarrow Z \qquad \text{strongly in } L^q(\Omega \times (0,T), W), \\ P_n Z &\longrightarrow Z \qquad \text{strongly in } L^q(\Omega \times (0,T), V). \end{aligned} \tag{3.3.51}$$

Step 2. Passing to the limit in the weak sense. From Lemma 3.3.5, we have

$$\mathbb{E} \sup_{t \in [0,T]} \|Y_n(t)\|_W^q \le C. \tag{3.3.52}$$

Then there exists a subsequence of Y_n, still denoted by Y_n such that

$$Y_n \rightharpoonup Y \qquad *\text{-weakly in } L^q(\Omega, L^\infty(0,T;W)). \tag{3.3.53}$$

Moreover, we have

$$P_n Y \longrightarrow Y \qquad \text{strongly in } L^q(\Omega \times (0,T), W). \tag{3.3.54}$$

Let us notice

$$|(S(y), \phi)| \le C\|y\|_W^3 \|\phi\|_2 \qquad \text{for any } y \in W \qquad \text{and } \phi \in H,$$

which implies that $S : W \to H^*$ and

$$\|S(y)\|_{H^*} \le C\|y\|_W^3, \qquad \forall y \in W.$$

Therefore

$$\|S(Y_n)\|^2_{L^2(\Omega, L^2(0,T;H^*))} = \mathbb{E}\int_0^T \|S(Y_n)\|^2_{H^*} \, ds \le C\mathbb{E} \sup_{t \in [0,T]} \|Y_n(t)\|_W^6 < C. \tag{3.3.55}$$

We also have

$$|(\operatorname{div} S(y),\phi)| \leq C\|y\|_W^3\|\phi\|_V \qquad \text{for any } \ y\in W \qquad \text{and } \ \phi\in V,$$

then

$$\|\operatorname{div} S(y)\|_{V^*} \leq C\|y\|_W^3, \qquad \forall y\in W,$$

and

$$\|\operatorname{div} S(Y_n)\|^2_{L^2(\Omega,L^2(0,T;W^*))} \leq \|\operatorname{div} S(Y_n)\|^2_{L^2(\Omega,L^2(0,T;V^*))} < C. \tag{3.3.56}$$

The operator N verifies

$$|(N(y),\phi)| \leq C\|y\|_W^2\|\phi\|_2 \qquad \text{for any } \ y\in W \qquad \text{and } \ \phi\in H,$$

In addition

$$|(\operatorname{div} N(y),\phi)| \leq C\|y\|_W^2\|\phi\|_W \qquad \text{for any } \ y,\phi\in W$$

which imply

$$\|N(Y_n)\|^2_{L^2(\Omega,L^2(0,T;H^*))} < C, \tag{3.3.57}$$

and

$$\|\operatorname{div} N(Y_n)\|^2_{L^2(\Omega,L^2(0,T;W^*))} \leq \|\operatorname{div} N(Y_n)\|^2_{L^2(\Omega,L^2(0,T;V^*))} < C. \tag{3.3.58}$$

Let us introduce the operator B, defined by

$$B(y) := -(y\cdot\nabla)y.$$

We have

$$|(B(y),\phi)| \leq C\|y\|_V^2\|\phi\|_V, \tag{3.3.59}$$

then

$$\|B(Y_n)\|^2_{L^2(\Omega,L^2(0,T;V^*))} \leq C_1\,\|Y_n\|^4_{L^4(\Omega,L^\infty(0,T;V))} < C. \tag{3.3.60}$$

The diffusion operator is bounded. Then there exist operators $N^*(t)$, $S^*(t)$, $B^*(t)$, $\sigma^*(t)$ and a subsequence on (n), that we still denote by (n), such that as $n\to\infty$ we have

$$\begin{aligned}
B(Y_n) &\rightharpoonup B^*(t) && \text{weakly in } \ L^2(\Omega\times(0,T),V^*),\\
N(Y_n) &\rightharpoonup N^*(t) && \text{weakly in } \ L^2(\Omega\times(0,T),H^*),\\
\operatorname{div} N(Y_n) &\rightharpoonup \operatorname{div} N^*(t) && \text{weakly in } \ L^2(\Omega\times(0,T),V^*),\\
S(Y_n) &\rightharpoonup S^*(t) && \text{weakly in } \ L^2(\Omega\times(0,T),H^*),\\
\operatorname{div} S(Y_n) &\rightharpoonup \operatorname{div} S^*(t) && \text{weakly in } \ L^2(\Omega\times(0,T),V^*),\\
\sigma(t,Y_n) &\rightharpoonup \sigma^*(t) && \text{weakly in } \ L^2(\Omega\times(0,T),(L^2(\mathbb{O}))^m).
\end{aligned} \tag{3.3.61}$$

Therefore, passing to the limit with respect to the weak topology, as $n\to\infty$, all terms in the equation (3.3.6), we derive that the limit function Y satisfies the stochastic differential equation

$$d\,(v(Y),\phi) = [(\nu\Delta Y + U,\phi) + \langle B^*(t),\phi\rangle + \langle \operatorname{div} N^*(t),\phi\rangle + \langle \operatorname{div} S^*(t),\phi\rangle]\,dt + (\sigma^*(t),\phi)\,d\mathcal{W}_t,$$
$$\forall\phi\in V. \tag{3.3.62}$$

Step 3. Passing to the limit in the strong sense up to a stopping time. Let us introduce the following convenient sequence (τ_M), $M \in \mathbb{N}$, of stopping times

$$\tau_M = \inf\{t \geq 0 : \|Y(t)\|_W \geq M\} \wedge T.$$

Proposition 3.3.6. *Let Y_n be the solution of (3.3.50) and P_nY the orthogonal projection of the weak limit Y on the space W_n. Then for M fixed we have*

$$\begin{aligned}
&\mathbb{E}\Big(\xi_2(t\wedge\tau_M)\|P_nY(t\wedge\tau_M)-Y_n(t\wedge\tau_M)\|_V^2\Big)+4\nu\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)\|D(P_nY-Y_n)\|_2^2 ds\\
&+\frac{\beta}{2}\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)\int_{\mathcal{O}}(|A_n|^2-|A|^2)^2 ds+\frac{\beta}{4}\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)\int_{\mathcal{O}}(|A_n|^2+|A|^2)|A(Y_n-Y)|^2 ds\\
&+\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)\|P_n\tilde{\sigma}-P_n\tilde{\sigma}^*\|_V^2 ds\to 0, \qquad as \quad n\to\infty, \qquad (3.3.63)
\end{aligned}$$

where

$$\xi_2(t)=e^{-D_3t-2D_4\int_0^t\|Y\|_W ds}$$

and D_3, D_4 are specific constants to be defined later on.

Proof. Taking the difference between equations (3.3.6) and (3.3.62), we write

$$\begin{aligned}
d(Y_n-P_nY,e_i)_V &= [(\nu\Delta(Y_n-Y),e_i)+\langle B(Y_n)-B^*(t),e_i\rangle]\,dt\\
&\quad+[\langle \operatorname{div} N(Y_n)-\operatorname{div} N^*(t),e_i\rangle+\langle \operatorname{div} S(Y_n)-\operatorname{div} S^*(t),e_i\rangle]\,dt\\
&\quad+(\sigma(t,Y_n)-\sigma^*(t),e_i)\,d\mathcal{W}_t, \qquad (3.3.64)
\end{aligned}$$

which holds for any $e_i \in W_n$, $i = 1, ..., n$.

The Itô's formula gives

$$\begin{aligned}
d(Y_n-P_nY,e_i)_V^2 &= 2(Y_n-P_nY,e_i)_V[(\nu\Delta(Y_n-Y),e_i)+\langle B(Y_n)-B^*(t),e_i\rangle]\,dt\\
&\quad+2(Y_n-P_nY,e_i)_V[\langle \operatorname{div} N(Y_n)-\operatorname{div} N^*(t),e_i\rangle+\langle \operatorname{div} S(Y_n)-\operatorname{div} S^*(t),e_i\rangle]\,dt\\
&\quad+2(Y_n-P_nY,e_i)_V(\sigma(t,Y_n)-\sigma^*(t),e_i)\,d\mathcal{W}_t+|(\sigma(t,Y_n)-\sigma^*(t),e_i)|^2\,dt.
\end{aligned}$$

Summing on $i = 1, \ldots, n$, we obtain

$$\begin{aligned}
&d\Big(\|Y_n-P_nY\|_V^2\Big)+4\nu\|D(Y_n-P_nY)\|_2^2 dt\\
&=2\nu(\Delta(P_nY-Y),Y_n-P_nY)\,dt+2\langle B(Y_n)-B^*(t),Y_n-P_nY\rangle dt\\
&+2[\langle \operatorname{div}(N(Y_n)-N^*(t)),Y_n-P_nY\rangle+\langle \operatorname{div}(S(Y_n)-S^*(t)),Y_n-P_nY\rangle]\,dt\\
&+\sum_{i=1}^n|(\sigma(t,Y_n)-\sigma^*(t),e_i)|^2\,dt+2(\sigma(t,Y_n)-\sigma^*(t),Y_n-P_nY)\,d\mathcal{W}_t. \qquad (3.3.65)
\end{aligned}$$

Now, we write each term in the right hand side of this equation in a convenient form

$$\begin{aligned}
&\langle \operatorname{div}(S(Y_n)-S^*(t)),Y_n-P_nY\rangle\\
&=\langle \operatorname{div}(S(Y_n)-S(Y)),Y_n-P_nY\rangle+\langle \operatorname{div}(S(Y)-S^*(t)),Y_n-P_nY\rangle\\
&=\langle \operatorname{div}(S(Y_n)-S(Y)),Y_n-Y\rangle+\langle \operatorname{div}(S(Y_n)-S(Y)),Y-P_nY\rangle\\
&+\langle \operatorname{div}(S(Y)-S^*(t)),Y_n-P_nY\rangle=g_n^1(t)+g_n^2(t)+g_n^3(t). \qquad (3.3.66)
\end{aligned}$$

Due to relation (3.2.12), we have

$$g_n^1(t) = -\frac{\beta}{4}\int_{\mathcal{O}}(|A_n|^2 - |A|^2)^2 - \frac{\beta}{4}\int_{\mathcal{O}}(|A_n|^2 + |A|^2)|A(Y_n - Y)|^2. \tag{3.3.67}$$

Using inequalities (3.3.66) and (3.3.67), equation (3.3.65) can be written as

$$\begin{aligned}
&d\left(\|Y_n - P_nY\|_V^2\right) + 4\nu\|D(Y_n - P_nY)\|_2^2 dt\\
&+\frac{\beta}{2}\int_{\mathcal{O}}(|A_n|^2 - |A|^2)^2 dt + \frac{\beta}{2}\int_{\mathcal{O}}(|A_n|^2 + |A|^2)|A(Y_n - Y)|^2 dt\\
&= 2\nu\,(\Delta(P_nY - Y), Y_n - P_nY)\,dt + 2\langle B(Y_n) - B^*(t), Y_n - P_nY\rangle dt\\
&+ 2\langle \operatorname{div}(N(Y_n) - N^*(t)), Y_n - P_nY\rangle dt + 2\left(g_n^2(t) + g_n^3(t)\right) dt\\
&+\sum_{i=1}^{n}|(\sigma(t,Y_n) - \sigma^*(t), e_i)|^2\,dt + 2\,(\sigma(t,Y_n) - \sigma^*(t), Y_n - P_nY)\,d\mathcal{W}_t.
\end{aligned} \tag{3.3.68}$$

We also have

$$\begin{aligned}
&\langle \operatorname{div}(N(Y_n) - N^*(t)), Y_n - P_nY\rangle\\
&= \langle \operatorname{div}(N(Y_n) - N(Y)), Y_n - Y\rangle + \langle \operatorname{div}(N(Y_n) - N(Y)), Y - P_nY\rangle\\
&+ \langle \operatorname{div}(N(Y) - N^*(t)), Y_n - P_nY\rangle = h_n^1(t) + h_n^2(t) + h_n^3(t).
\end{aligned} \tag{3.3.69}$$

Applying Lemma 3.2.6 with $3\epsilon = \frac{\beta}{8}$, we have

$$\begin{aligned}
h_n^1(t) \le{}& \frac{\beta}{8}\int_{\mathcal{O}}|A(Y_n - Y)|^2\left(|A|^2 + |A_n|^2\right) + C_1\|Y_n - Y\|_V^2\\
&+\frac{C}{1-\lambda}\epsilon^{\frac{\lambda-1}{\lambda+3}}\|Y_n - P_nY\|_{H^1}^{\frac{4(\lambda+1)}{\lambda+3}}\|Y\|_{H^2}^{\frac{4}{\lambda+3}} + \frac{C}{1-\lambda}\epsilon^{\frac{\lambda-1}{\lambda+3}}\|P_nY - Y\|_{H^1}^{\frac{4(\lambda+1)}{\lambda+3}}\|Y\|_{H^2}^{\frac{4}{\lambda+3}}
\end{aligned} \tag{3.3.70}$$

for any $\lambda \in]0,1[$. Let us set

$$h_n^4(t) = \frac{C}{1-\lambda}\epsilon^{\frac{\lambda-1}{\lambda+3}}\|P_nY - Y\|_{H^1}^{\frac{4(\lambda+1)}{\lambda+3}}\|Y\|_{H^2}^{\frac{4}{\lambda+3}}.$$

Proceeding analogously with the convective term, we deduce

$$\begin{aligned}
&\langle B(Y_n) - B^*(t), Y_n - P_nY\rangle\\
&= \langle B(Y_n) - B(Y), Y_n - Y\rangle + \langle B(Y_n) - B(Y), Y - P_nY\rangle\\
&+ \langle B(Y) - B^*(t), Y_n - P_nY\rangle = b_n^1(t) + b_n^2(t) + b_n^3(t).
\end{aligned} \tag{3.3.71}$$

In addition

$$|b_n^1(t)| \le C_2\|Y\|_W\|Y_n - Y\|_V^2. \tag{3.3.72}$$

Denoting by $\tilde{\sigma}_n$, $\tilde{\sigma}$ and $\tilde{\sigma}^*$ the solutions of the Stokes system (3.2.1) for $f = \sigma(t,Y_n)$, $f = \sigma(t,Y)$ and $f = \sigma^*(t)$, respectively, we have

$$(\sigma(t,Y_n) - \sigma^*(t), e_i) = (\tilde{\sigma}_n - \tilde{\sigma}^*, e_i)_V, \qquad i = 1,2,\ldots,n.$$

Then

$$\sum_{i=1}^{n} |(\sigma(t,Y_n)-\sigma^*(t),e_i)|^2 = \|P_n\tilde{\sigma}_n - P_n\tilde{\sigma}^*\|_V^2.$$

The standard relation $x^2 = (x-y)^2 - y^2 + 2xy$ allows to write

$$\begin{aligned}\|P_n\tilde{\sigma}_n - P_n\tilde{\sigma}^*\|_V^2 &= \|P_n\tilde{\sigma}_n - P_n\tilde{\sigma}\|_V^2 - \|P_n\tilde{\sigma} - P_n\tilde{\sigma}^*\|_V^2 \\ &\quad + 2(P_n\tilde{\sigma}_n - P_n\tilde{\sigma}^*, P_n\tilde{\sigma} - P_n\tilde{\sigma}^*)_V.\end{aligned}$$

From the properties of the solutions of the Stokes system (3.2.1) and (3.3.3), we have

$$\|P_n\tilde{\sigma}_n - P_n\tilde{\sigma}\|_V^2 \le \|\tilde{\sigma}_n - \tilde{\sigma}\|_V^2 \le \|\sigma(t,Y_n)-\sigma(t,Y)\|_2^2 \le K\|Y_n - Y\|_V^2,$$

then

$$\begin{aligned}\|P_n\tilde{\sigma}_n - P_n\tilde{\sigma}^*\|_V^2 &\le K\|Y_n - Y\|_V^2 - \|P_n\tilde{\sigma} - P_n\tilde{\sigma}^*\|_V^2 \\ &\quad + 2(P_n\tilde{\sigma}_n - P_n\tilde{\sigma}^*, P_n\tilde{\sigma} - P_n\tilde{\sigma}^*)_V \\ &\le 2K\|Y_n - P_nY\|_V^2 + C\|P_nY - Y\|_V^2 - \|P_n\tilde{\sigma} - P_n\tilde{\sigma}^*\|_V^2 \\ &\quad + 2(P_n\tilde{\sigma}_n - P_n\tilde{\sigma}^*, P_n\tilde{\sigma} - P_n\tilde{\sigma}^*)_V. \end{aligned} \tag{3.3.73}$$

Let us set $D_3 := 2(K+2C_1)$ and $D_4 := 2C_2$. The positive constants K, C_1 and C_2 in (3.3.3), (3.3.70) and (3.3.72) are independent of n.

We introduce the auxiliary function

$$\xi_2(t) = e^{-D_3 t - 2D_4\int_0^t \|Y\|_W ds}.$$

Now, applying the Itô formula and using the equality (3.3.68), we get

$$\begin{aligned}&d\left(\xi_2(t)\|Y_n - P_nY\|_V^2\right) + 4\nu\xi_2(t)\|D(Y_n - P_nY)\|_2^2\,dt \\ &+ \frac{\beta}{2}\xi_2(t)\||A_n|^2 - |A|^2\|_2^2\,dt + \frac{\beta}{2}\xi_2(t)\|\sqrt{|A_n|^2+|A|^2}\,|A(Y_n - Y)|\|_2^2\,dt \\ &= 2\nu\xi_2(t)(\Delta(P_nY - Y), Y_n - P_nY)\,dt \\ &+ 2\xi_2(t)\langle B(Y_n) - B^*(t), Y_n - P_nY\rangle\,dt \\ &+ 2\xi_2(t)\langle \mathrm{div}N(Y_n) - \mathrm{div}N^*(t), Y_n - P_nY\rangle\,dt + 2\xi_2(t)\left(g_n^2(t) + g_n^3(t)\right)dt \\ &+ \xi_2(t)\sum_{i=1}^{n}|(\sigma(t,Y_n) - \sigma^*(t), e_i)|^2\,dt \\ &+ 2\xi_2(t)(\sigma(t,Y_n) - \sigma^*(t), Y_n - P_nY)\,d\mathcal{W}_t \\ &- D_3\xi_2(t)\|P_nY - Y_n\|_V^2\,dt - 2D_4\xi_2(t)\|Y\|_W\|Y_n - P_nY\|_V^2\,dt.\end{aligned}$$

Incorporate in this equation the relations (3.3.69), (3.3.70), (3.3.71), (3.3.72) and (3.3.73), we deduce

$$
\begin{aligned}
&d\Big(\xi_2(t)\|Y_n-P_nY\|_V^2\Big)+4\nu\xi_2(t)\|D(Y_n-P_nY)\|_2^2\,dt+\frac{\beta}{2}\xi_2(t)\||A_n|^2-|A|^2\|_2^2\,dt\\
&+\frac{\beta}{4}\xi_2(t)\|\sqrt{|A_n|^2+|A|^2}\,|A(Y_n-Y)|\|_2^2\,dt+\xi_2(t)\|P_n\tilde{\sigma}-P_n\tilde{\sigma}^*\|_V^2\,dt\\
&\leq 2\nu\xi_2(t)(\Delta(P_nY-Y),Y_n-P_nY)\,dt+\xi_2(t)\frac{2C}{1-\lambda}\epsilon^{\frac{\lambda-1}{\lambda+3}}\|P_nY-Y_n\|_{H^1}^{\frac{4(\lambda+1)}{\lambda+3}}\|Y\|_{H^2}^{\frac{4}{\lambda+3}}\,dt\\
&+2\xi_2(t)\Big[b_n^2(t)+b_n^3(t)+h_n^2(t)+h_n^3(t)+h_n^4(t)+g_n^2(t)+g_n^3(t)\Big]\,dt\\
&+\xi_2(t)\Big[C(1+\|Y\|_W)\|P_nY-Y\|_V^2+2(P_n\tilde{\sigma}_n-P_n\tilde{\sigma}^*,P_n\tilde{\sigma}-P_n\tilde{\sigma}^*)_V\Big]\,dt\\
&+2\xi_2(t)\,(\sigma(t,Y_n)-\sigma^*(t),Y_n-P_nY)\,d\mathscr{W}_t.
\end{aligned}
$$

Integrating over the time interval $(0,t\wedge\tau_M)$, $t\in[0,T]$, and taking the expectation, we derive

$$
\begin{aligned}
&\mathbb{E}\Big(\xi_2(t\wedge\tau_M)\|P_nY(t\wedge\tau_M)-Y_n(t\wedge\tau_M)\|_V^2\Big)+4\nu\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)\|D(P_nY-Y_n)\|_2^2\,ds\\
&+\frac{\beta}{2}\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)\||A_n|^2-|A|^2\|_2^2\,ds+\frac{\beta}{4}\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)\|\sqrt{|A_n|^2+|A|^2}\,|A(Y_n-Y)|\|_2^2ds\\
&+\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)\|P_n\tilde{\sigma}-P_n\tilde{\sigma}^*\|_V^2\,ds \qquad (3.3.74)\\
&\leq 2\nu\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)(\Delta(Y-P_nY),P_nY-Y_n)\,ds\\
&+2\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)\Big[b_n^2(s)+b_n^3(s)+h_n^2(s)+h_n^3(s)+h_n^4(s)+g_n^2(s)+g_n^3(s)\Big]\,ds\\
&+\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)\Big[C(1+M)\|P_nY-Y\|_V^2+2(P_n\tilde{\sigma}_n-P_n\tilde{\sigma}^*,P_n\tilde{\sigma}-P_n\tilde{\sigma}^*)_V\Big]\,ds\\
&+\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)\frac{2C}{1-\lambda}\epsilon^{\frac{\lambda-1}{\lambda+3}}\|P_nY-Y_n\|_{H^1}^{\frac{4(\lambda+1)}{\lambda+3}}\|Y\|_{H^2}^{\frac{4}{\lambda+3}}\,ds\\
&=J_n^1+J_n^2+J_n^3+J_n^4. \qquad (3.3.75)
\end{aligned}
$$

Here, we assume that

$$
r_n(t)=J_n^1+J_n^2+J_n^3\to 0. \qquad (3.3.76)
$$

This result will be proved in a lemma at the end of this proposition.

Let us define

$$
\begin{aligned}
a_n(t)=&\mathbb{E}\Big(\xi_2(t\wedge\tau_M)\|P_nY(t\wedge\tau_M)-Y_n(t\wedge\tau_M)\|_V^2\Big)+4\nu\mathbb{E}\int_0^t 1_{[0,\tau_M]}(s)\xi_2(s)\|D(P_nY-Y_n)\|_2^2\,ds\\
&+\frac{\beta}{2}\mathbb{E}\int_0^t 1_{[0,\tau_M]}(s)\xi_2(s)\||A_n|^2-|A|^2\|_2^2\,ds\\
&+\frac{\beta}{4}\mathbb{E}\int_0^t 1_{[0,\tau_M]}(s)\xi_2(s)\|\sqrt{|A_n|^2+|A|^2}\,|A(Y_n-Y)|\|_2^2ds\\
&+\mathbb{E}\int_0^t 1_{[0,\tau_M]}(s)\xi_2(s)\|P_n\tilde{\sigma}-P_n\tilde{\sigma}^*\|_V^2\,ds. \qquad (3.3.77)
\end{aligned}
$$

Taking into account (3.3.74), (3.3.76) and the concavity of the function $x \to x^{\frac{2(\lambda+1)}{\lambda+3}}$, $\lambda \in [0,1]$, we derive

$$\begin{aligned}
a_n(t) &\le r_n(t) + \frac{2C}{1-\lambda}\mathbb{E}\int_0^t 1_{[0,\tau_M]}(s)\xi_2(s)\epsilon^{\frac{\lambda-1}{\lambda+3}}\|P_nY - Y_n\|_{H^1}^{\frac{4(\lambda+1)}{\lambda+3}}\|Y\|_{H^2}^{\frac{4}{\lambda+3}}ds \\
&\le r_n(t) + M^{\frac{4}{\lambda+3}}\frac{2C}{1-\lambda}\epsilon^{\frac{\lambda-1}{\lambda+3}}\mathbb{E}\int_0^t \xi_2(s\wedge\tau_M)\|P_nY(s\wedge\tau_M) - Y_n(s\wedge\tau_M)\|_{H^1}^{\frac{4(\lambda+1)}{\lambda+3}}ds \\
&\le r_n(t) + M^{\frac{4}{\lambda+3}}\frac{2C}{1-\lambda}\epsilon^{\frac{\lambda-1}{\lambda+3}}\mathbb{E}\int_0^t \left(\xi_2(s\wedge\tau_M)\|P_nY(s\wedge\tau_M) - Y_n(s\wedge\tau_M)\|_V^2\right)^{\frac{2(\lambda+1)}{\lambda+3}}ds \\
&\le r_n(t) + M^{\frac{4}{\lambda+3}}\frac{2C}{1-\lambda}\epsilon^{\frac{\lambda-1}{\lambda+3}}\left(\int_0^t \mathbb{E}\xi_2(s\wedge\tau_M)\|P_nY(s\wedge\tau_M) - Y_n(s\wedge\tau_M)\|_V^2ds\right)^{\frac{2(\lambda+1)}{\lambda+3}} \\
&\le r_n(t) + M^{\frac{4}{\lambda+3}}\frac{2C}{1-\lambda}\epsilon^{\frac{\lambda-1}{\lambda+3}}\left(\int_0^t a_n(s)ds\right)^{\frac{2(\lambda+1)}{\lambda+3}},
\end{aligned} \tag{3.3.78}$$

which yields

$$\limsup_{n\to\infty} a_n(t) \le \limsup_{n\to\infty} r_n(t) + M^{\frac{4}{\lambda+3}}\frac{2C}{1-\lambda}\epsilon^{\frac{\lambda-1}{\lambda+3}}\left(\int_0^t \limsup_{n\to\infty} a_n(s)ds\right)^{\frac{2(\lambda+1)}{\lambda+3}}. \tag{3.3.79}$$

Denoting

$$f(t) := \int_0^t \limsup_{n\to\infty} a_n(s)ds,$$

and knowing that $\lim_{n\to\infty} r_n(t) = 0$, (3.3.79) can be written as

$$f'(t) \le M^{\frac{4}{\lambda+3}}\frac{2C}{1-\lambda}\epsilon^{\frac{\lambda-1}{\lambda+3}}(f(t))^{\frac{4(\lambda+1)}{\lambda+3}}. \tag{3.3.80}$$

Here, we can proceed as in [10] in order to verify that $f \equiv 0$. Since $f(0) = 0$ and

$$\left((f(t))^{\frac{1-\lambda}{\lambda+3}}\right)' \le \frac{2C}{\lambda+3}M^{\frac{4}{\lambda+3}}\epsilon^{\frac{\lambda+1}{\lambda+3}},$$

we have

$$f(t) \le \left(\frac{2C}{\lambda+3}M^{\frac{4}{\lambda+3}}\epsilon^{\frac{\lambda+1}{\lambda+3}}t\right)^{\frac{\lambda+3}{1-\lambda}}.$$

Considering $T_0 = \frac{3}{4CM^{\frac{4}{3}}\epsilon}$, we have $\frac{2C}{\lambda+3}M^{\frac{4}{\lambda+3}}\epsilon^{\frac{\lambda+1}{\lambda+3}}t \le \frac{1}{2}$. Taking $\lambda \to 1$, we get $f(t) = 0$, $\forall t \in [0, T_0]$. By an extension argument, we obtain $f(t) = 0$, $\forall t \in [0, T]$.

$$\begin{aligned}
&\mathbb{E}\left(\xi_2(t\wedge\tau_M)\|P_nY(t\wedge\tau_M) - Y_n(t\wedge\tau_M)\|_V^2\right) + 4\nu\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)\|D(P_nY - Y_n)\|_2^2ds \\
&+\frac{\beta}{2}\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)\int_{\mathcal{O}}(|A_n|^2 - |A|^2)^2ds + \frac{\beta}{4}\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)\int_{\mathcal{O}}(|A_n|^2 + |A|^2)|A(Y_n - Y)|^2ds \\
&+\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)\|P_n\tilde{\sigma} - P_n\tilde{\sigma}\|_V^2ds \to 0, \qquad \text{as} \quad n \to \infty.
\end{aligned}$$

■

Lemma 3.3.7. *Let $J_n^1(t)$, $J_n^2(t)$, $J_n^3(t)$ be the terms introduced in (3.3.74). Then for all $t \in [0,T]$, $J_n^i(t) \to 0$, for $i = 1,2,3$.*

Proof. Using (3.3.52)-(3.3.53) and the properties of the projection P_n, we have

$$\begin{aligned}
|J_n^1(t)| &= \Big|2\nu\mathbb{E}\int_0^t \xi_2(s)(1_{[0,\tau_M]}(s)\Delta(Y - P_nY), P_nY - Y_n)\,ds\Big| \\
&\leq C\|P_nY - Y\|_{L^2(\Omega\times(0,t),H^2)}\|P_nY - Y_n\|_{L^2(\Omega\times(0,t),W)} \\
&\leq C\|P_nY - Y\|_{L^2(\Omega\times(0,T),W)}\Big(\|Y\|_{L^2(\Omega\times(0,T),W)} + \|Y_n\|_{L^2(\Omega\times(0,T),H^2)}\Big) \\
&\leq C\|P_nY - Y\|_{L^2(\Omega\times(0,T),H^2)}
\end{aligned}$$

which goes to zero, as $n \to \infty$, by (3.3.54).

$$J_n^2(t) = 2\mathbb{E}\int_0^{t\wedge\tau_M} \xi_2(s)\Big[b_n^2(s) + b_n^3(s) + h_n^2(s) + h_n^3(s) + h_n^4(s) + g_n^2(s) + g_n^3(s)\Big]\,ds$$

From (3.3.59), (3.3.35) and (3.3.54), we deduce

$$\begin{aligned}
\left|2\mathbb{E}\int_0^{t\wedge\tau_M} \xi_2(s)b_n^2(s)\right|ds &= \left|2\mathbb{E}\int_0^{t\wedge\tau_M} \xi_2(s)\langle B(Y_n) - B(Y), P_nY - Y\rangle\right| \\
&\leq C\mathbb{E}\sup_{t\in[0,T]}\|Y_n\|_V^4\;\mathbb{E}\|P_nY - Y\|^2_{L^2(0,T;V)} \to 0, \quad \text{as} \quad n \to \infty.
\end{aligned}$$

Convergences (3.3.53) and (3.3.54) give that

$$P_nY - Y_n \to 0 \qquad \text{weakly in } L^2(\Omega\times(0,T),W),$$

then for any operator $R \in L^2(\Omega\times(0,T),W^*)$ we have

$$\mathbb{E}\int_0^T \langle R, P_nY - Y_n\rangle\,ds \to 0, \qquad \text{as } n \to \infty.$$

The function $1_{[0,\tau_M]}(s)\xi_2(s)$ is bounded, then

$$\begin{aligned}
&\|1_{[0,\tau_M]}(s)\xi_2(s)(B(Y) - B^*)\|^2_{L^2(\Omega\times(0,T),W^*)} \\
\leq\;& C\Big(\|B(Y)\|^2_{L^2(\Omega\times(0,T),W^*)} + \|B^*\|^2_{L^2(\Omega\times(0,T),W^*)}\Big) \leq C,
\end{aligned}$$

by (3.3.52), (3.3.60) and (3.3.61). Therefore, as $n \to \infty$, we have

$$2\mathbb{E}\int_0^{t\wedge\tau_M} \xi_2(s)b_n^3(s) = 2\mathbb{E}\int_0^t \langle 1_{[0,\tau_M]}(s)\xi_2(s)(B(Y) - B^*(s)), P_nY - Y_n\rangle\,ds \to 0.$$

Using the same reasoning, we show that

$$2\mathbb{E}\int_0^{t\wedge\tau_M} \xi_2(s)h_n^2(s) \to 0, \quad 2\mathbb{E}\int_0^{t\wedge\tau_M} \xi_2(s)h_n^3(s) \to 0.$$

By the definition of the stopping time τ_M, we have $1_{[0,\tau_M]}(s)\xi_2(s)\|Y\|_W^{\frac{4}{\lambda+3}} \le M^{\frac{4}{\lambda+3}}$, so

$$\begin{aligned}\left|2\mathbb{E}\int_0^{t\wedge\tau_M}\xi_2(s)h_n^4(s)\right| &\le \frac{C}{1-\lambda}\epsilon^{\frac{\lambda-1}{\lambda+3}}\left|2\mathbb{E}\int_0^T 1_{[0,\tau_M]}(s)\xi_2(s)\|P_nY-Y\|_V^{\frac{4(\lambda+1)}{\lambda+3}}\|Y\|_W^{\frac{4}{\lambda+3}}\right| \\ &\le \frac{C}{1-\lambda}\epsilon^{\frac{\lambda-1}{\lambda+3}}M^{\frac{4}{\lambda+3}}\left|2\mathbb{E}\int_0^T\|P_nY-Y\|_V^{\frac{4(\lambda+1)}{\lambda+3}}\right| \\ &\le C(M,\lambda)\|P_nY-Y\|_{L^2(\Omega\times(0,T),V)}^{\frac{2(\lambda+1)}{\lambda+3}} \to 0\end{aligned}$$

Similarly we verify that the remaining terms in $J_n^2(t)$ converges to 0, as well as $J_n^3(t)$ converges to 0, as $n\to\infty$.

■

From (3.3.63), the following strong convergences hold

$$\lim_{n\to\infty}\mathbb{E}\Big(\xi_2(\tau_M)\|P_nY(\tau_M)-Y_n(\tau_M)\|_V^2\Big)=0, \tag{3.3.81}$$

$$\lim_{n\to\infty}\mathbb{E}\int_0^{\tau_M}\xi_2(s)\|D(P_nY-Y_n)\|_2^2ds=0, \tag{3.3.82}$$

$$\lim_{n\to\infty}\mathbb{E}\int_0^{\tau_M}\xi_2(s)\|\,|A_n|^2-|A|^2\|_2^2\,ds=0, \tag{3.3.83}$$

$$\lim_{n\to\infty}\mathbb{E}\int_0^{\tau_M}\xi_2(s)\|\sqrt{|A_n|^2+|A|^2}\,|A(Y_n-Y)|\|_2^2ds=0, \tag{3.3.84}$$

$$\lim_{n\to\infty}\mathbb{E}\int_0^{\tau_M}\xi_2(s)\|P_n\tilde{\sigma}-P_n\tilde{\sigma}^*\|_V^2ds=0, \tag{3.3.85}$$

for each $M\in\mathbb{N}$. Since there exists a strictly positive constant μ, such that $\mu\le 1_{[0,\tau_M]}(s)\xi_2(s)\le 1$, it follows from the Korn inequality (3.1.9) and (3.3.54) that

$$\lim_{n\to\infty}\mathbb{E}\int_0^{\tau_M}\xi_2(s)\|D(P_nY-Y_n)\|_2^2ds=0 \quad\text{implies}\quad \lim_{n\to\infty}\mathbb{E}\int_0^{\tau_M}\|Y-Y_n\|_V^2ds=0. \tag{3.3.86}$$

In addition, we have

$$\lim_{n\to\infty}\mathbb{E}\int_0^{\tau_M}\|\,|A_n|^2-|A|^2\|_2^2\,ds=0, \tag{3.3.87}$$

$$\lim_{n\to\infty}\mathbb{E}\int_0^{\tau_M}\|\sqrt{|A_n|^2+|A|^2}\,|A(Y_n-Y)|\|_2^2ds=0. \tag{3.3.88}$$

Considering (3.3.51), we also derive

$$\mathbb{E}\int_0^{\tau_M}\|\tilde{\sigma}-\tilde{\sigma}^*\|_V^2ds=0. \tag{3.3.89}$$

Step 4. Identification of $B^*(t)$ *with* $B(Y)$, $\operatorname{div}N^*(t)$ *with* $\operatorname{div}N(Y)$, $\operatorname{div}S^*(t)$ *with* $\operatorname{div}S(Y)$ *and* $\sigma^*(t)$ *with* $\sigma(t,Y)$ *on* $[0,\tau_M]$ *for each* M.

Now, we are able to show that the limit function Y satisfies equation (3.3.1). Integrating equation (3.3.62) on the time interval $(0, \tau_M \wedge t)$, we derive

$$\begin{aligned}(\upsilon(Y(\tau_M \wedge t)), \phi) - (\upsilon(Y_0), \phi) \quad = \quad & \int_0^{\tau_M \wedge t} \Big[(\nu \Delta Y + U, \phi) + \langle B^*(s), \phi\rangle + \langle \operatorname{div} N^*(s), \phi\rangle \\ & + \langle \operatorname{div} S^*(s), \phi\rangle\Big] ds + \int_0^{\tau_M \wedge t} (\sigma^*(s), \phi)\, d\mathcal{W}_s \qquad (3.3.90)\end{aligned}$$

for any $\phi \in V$. From (3.3.89) it follows that

$$1_{[0,\tau_M]}(t)\tilde{\sigma} = 1_{[0,\tau_M]}(t)\tilde{\sigma}^* \qquad \text{a.e. in } \Omega \times (0,T),$$

which implies

$$1_{[0,\tau_M]}(t)\sigma(t,Y) = 1_{[0,\tau_M]}(t)\sigma^*(t) \qquad \text{a. e. in } \Omega \times (0,T) \qquad (3.3.91)$$

by (3.2.1). Since $B(Y_n) - B(Y) = (Y_n \cdot \nabla)(Y_n - Y) + (Y_n - Y) \cdot \nabla Y$, we verify that

$$\|B(Y_n) - B(Y)\|_{V^*} \leq C\,(\|Y_n\|_V + \|Y\|_V)\,\|Y_n - Y\|_V.$$

Then for any $\varphi \in L^\infty(\Omega \times (0,T), V)$, using (3.3.52), (3.3.53)

$$\begin{aligned}&\Big|\mathbb{E}\int_0^T 1_{[0,\tau_M]}(s)\langle B(Y_n) - B(Y),\ \varphi\rangle\, ds\Big| \\ &\qquad \leq C\mathbb{E}\int_0^T 1_{[0,\tau_M]}(s)\,(\|Y_n\|_V + \|Y\|_V)\,\|Y_n - Y\|_V \|\varphi\|_V\, ds \\ &\qquad \leq C\|\varphi\|_{L^\infty(\Omega\times(0,T),V)}\mathbb{E}\int_0^{\tau_M} (\|Y_n\|_V + \|Y\|_V)\,\|Y_n - Y\|_V\, ds \\ &\qquad \leq C\|\varphi\|_{L^\infty(\Omega\times(0,T),V)}\left(\mathbb{E}\int_0^{\tau_M} \|Y_n - Y\|_V^2\, ds\right)^{\frac{1}{2}} \to 0, \qquad \text{as } n \to \infty.\end{aligned}$$

Taking into account $(3.3.61)_1$ and that the space $L^\infty(\Omega \times (0,T), V)$ is dense in $L^2(\Omega \times (0,T), V)$, we obtain

$$1_{[0,\tau_M]}(s)B^*(s) = 1_{[0,\tau_M]}(s)B(Y) \qquad \text{a. e. in } \Omega \times (0,T). \qquad (3.3.92)$$

From (3.2.20), we have

$$\begin{aligned}\left|\langle \operatorname{div}(N(Y_n) - N(Y)), \phi\rangle\right| \leq\ & C\epsilon \left\|A(Y_n - Y)\sqrt{|A_n|^2 + |A|^2}\right\|_2 \|\phi\|_V \\ & + C\|Y_n - Y\|_V\,(\|Y_n\|_W + \|Y\|_W)\,\|\phi\|_W. \qquad (3.3.93)\end{aligned}$$

Then for any $\phi \in L^\infty(\Omega\times(0,T),W)$, using (3.3.86) and (3.3.88), we deduce

$$\begin{aligned}
&\left|\mathbb{E}\int_0^T 1_{[0,\tau_M]}(s)\langle \operatorname{div}(N(Y_n)-N(Y)),\phi\rangle\, ds\right| \\
&\leq C\mathbb{E}\int_0^T 1_{[0,\tau_M]}(s)\left\|A(Y_n-Y)\sqrt{|A_n|^2+|A|^2}\right\|_2 \|\phi\|_V\, ds \\
&+ C\mathbb{E}\int_0^T 1_{[0,\tau_M]}(s)\|Y_n-Y\|_V\,(\|Y_n\|_W+\|Y\|_W)\,\|\phi\|_W\, ds \\
&\leq C\|\phi\|_{L^\infty(\Omega\times(0,T),W)}\mathbb{E}\int_0^{\tau_M}\left\|A(Y_n-Y)\sqrt{|A_n|^2+|A|^2}\right\|_2 ds \\
&+ C\|\phi\|_{L^\infty(\Omega\times(0,T),W)}\mathbb{E}\int_0^{\tau_M}(\|Y_n\|_W+\|Y\|_W)\|Y_n-Y\|_V\, ds \\
&\leq C\|\phi\|_{L^\infty(\Omega\times(0,T),W)}\mathbb{E}\int_0^{\tau_M}\left\|A(Y_n-Y)\sqrt{|A_n|^2+|A|^2}\right\|_2^2 ds \\
&+ C\|\phi\|_{L^\infty(\Omega\times(0,T),W)}\left(\mathbb{E}\int_0^{\tau_M}\|Y_n-Y\|_V^2\, ds\right)^{\frac{1}{2}} \to 0, \qquad \text{as } n\to\infty.
\end{aligned}$$

Therefore

$$1_{[0,\tau_M]}(s)\operatorname{div}(N^*(s)) = 1_{[0,\tau_M]}(s)\operatorname{div}(N(Y)) \qquad \text{a.e. in } \Omega\times(0,T). \tag{3.3.94}$$

Using the same reasoning, we show

$$1_{[0,\tau_M]}(s)\operatorname{div}(S^*(s)) = 1_{[0,\tau_M]}(s)\operatorname{div}(S(Y)) \qquad \text{a.e. in } \Omega\times(0,T). \tag{3.3.95}$$

Namely, from (3.2.12), we have

$$\left|\langle \operatorname{div}(S(Y)-S(Y_n)),\phi\rangle\right| \leq C\|Y\|_W^2\|Y-Y_n\|_V\|\phi\|_W + C\|Y_n\|_W\left\||A|^2-|A_n|^2\right\|_2\|\phi\|_W, \tag{3.3.96}$$

and (3.3.86) and (3.3.87) gives

$$\begin{aligned}
&\left|\mathbb{E}\int_0^T 1_{[0,\tau_M]}(s)\langle \operatorname{div}(S(Y)-S(Y_n),\phi\rangle\, ds\right| \\
&\leq C\mathbb{E}\int_0^T 1_{[0,\tau_M]}(s)\|Y\|_W^2\|Y-Y_n\|_V\|\phi\|_W\, ds \\
&+ C\mathbb{E}\int_0^T 1_{[0,\tau_M]}(s)\|Y_n\|_W\left\||A|^2-|A_n|^2\right\|_2\|\phi\|_W\, ds \\
&\leq C\|\phi\|_{L^\infty(\Omega\times(0,T),W)}\mathbb{E}\int_0^{\tau_M}\|Y\|_W^2\|Y_n-Y\|_V\, ds \\
&+ C\|\phi\|_{L^\infty(\Omega\times(0,T),W)}\mathbb{E}\int_0^{\tau_M}\|Y_n\|_W\left\||A|^2-|A_n|^2\right\|_2 \\
&\leq C(M)\|\phi\|_{L^\infty(\Omega\times(0,T),W)}\mathbb{E}\int_0^{\tau_M}\|Y_n-Y\|_V\, ds \\
&+ C(M)\|\phi\|_{L^\infty(\Omega\times(0,T),W)}\mathbb{E}\int_0^{\tau_M}\left\||A|^2-|A_n|^2\right\|_2 \to 0, \qquad \text{as } n\to\infty.
\end{aligned}$$

By introducing identities (3.3.91), (3.3.92), (3.3.94) and (3.3.95) in equation (3.3.90), it follows that

$$(v(Y(\tau_M \wedge t)), \phi) - (v(Y_0), \phi) = \int_0^{\tau_M \wedge t} \Big[(\nu \Delta Y + U, \phi) + \langle B(Y) + \operatorname{div}(N(Y)) + \operatorname{div}(S(Y)), \phi \rangle \Big] ds + \int_0^{\tau_M \wedge t} (\sigma(s, Y), \phi) \, d\mathcal{W}_s. \tag{3.3.97}$$

Reasoning as in (3.3.25) we have $\tau_M \to T$ a.e. in Ω, as $M \to \infty$. We can pass to the limit in each term of equation (3.3.97) in $L^1(\Omega \times (0,T))$, as $M \to \infty$, by applying the Lebesgue dominated convergence theorem and the Burkholder-Davis-Gundy inequality for the last (stochastic) term, deriving an equivalent formulation of equation (3.3.1) a.e. in $\Omega \times (0,T)$.

Step 5. Uniqueness. In order to prove uniqueness, we take two solutions Y_1 and Y_2, and consider the difference $Y = Y_1 - Y_2$. Using similar arguments as in the previous steps, introducing the function

$$\xi_3(t) = e^{-\frac{1}{2} D_3 t - D_4 \int_0^t \|Y_1\|_W ds},$$

we show that

$$\begin{aligned} \mathbb{E}\Big(\xi_3(t) \|Y(t)\|_V^2\Big) &+ 4\nu \mathbb{E} \int_0^t \xi_3(s) \|D(Y)\|_2^2 ds \\ &+ \frac{\beta}{2} \mathbb{E} \int_0^t \xi_3(s) \int_{\mathcal{O}} (|A(Y_1)|^2 - |A(Y_2)|^2)^2 ds \\ &+ \frac{\beta}{4} \mathbb{E} \int_0^t \xi_3(s) \int_{\mathcal{O}} (|A(Y_1)|^2 + |A(Y_2)|^2) |A(Y)|^2 ds \\ &= 0 \quad \text{for a.e. } t \in [0,T]. \end{aligned}$$

Therefore, for a.e. $t \in [0,T]$, we have

$$\mathbb{E}\Big(\xi_3(t) \|Y(t)\|_V^2\Big) = 0.$$

Since ξ_3 is a positive function, we deduce that for a.e. $t \in [0,T]$

$$Y_1(t) = Y_2(t), \quad P-\text{a.s.}.$$

■

4

Gaussian fields and approximation by Gaussian Markov random fields

In this chapter we consider the SPDE

$$(k^2 - \Delta)^{\alpha/2} \tau x = \mathcal{W} \, ,$$

where $\mathcal{W}$ is a Gaussian white noise. The solution can be deduced by applying the Fourier methods. In addition, it can be verified that the solution is a Gaussian field with Matérn covariance. In spatial statistics, Matérn covariance plays an important role, since it describes quite well the behaviour of several spatial phenomena, such as epidemics, rainfall, social sciences, etc. To present a self-contained text we first solve SPDE and then we use the FEM to obtain a finite representation $\tilde{x}$ of the solution of the SPDE,

$$\tilde{x}(s) = \sum_{i=1}^{m} x_i \varphi_i(s) \, ,$$

where $\mathbf{x} = (x_1, \cdots, x_m)$ is a Gaussian random vector of weights, and $\{\varphi_i : i = 1, \dots m\}$ is a set of finite element basis functions. We will verify that $\mathbf{x}$ is a Gaussian Markov random field, and explicitly calculate matrices $\mathbf{C}$ and $\mathbf{G}$, needed to obtain the precision matrix $\mathbf{Q}$ of $\mathbf{x}$. These results will used in Chapter 5.

Definition 4.0.1. *Consider $D \subset \mathbb{R}^d$ and $s \in D$. We say that $\{x(s) : s \in D \subset \mathbb{R}^d\}$, $x(s)$ for short, is a continuously indexed Gaussian field if all finite collections $\{x(s_i) : i = 1, \dots, n\}$ are jointly Gaussian distributed.*

It should be mentioned that a Gaussian field is perfectly defined by its mean and covariance functions.

Next we present the *Fourier Transform* for $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. In order to apply the Fourier transform techniques to solve our stochastic partial differential equation, we extend the definition to generalized functions and to generalized random functions.

Definition 4.0.2. *Consider $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. The* Fourier transform *of f is defined as*

$$\hat{f}(\lambda) \equiv \mathcal{F}\{f(s)\}(\lambda) = \int_{\mathbb{R}^d} f(s) e^{-is\lambda} \, ds$$

and the inverse Fourier *Transform is defined as*

$$\check{f}(s) \equiv \mathscr{F}^{-1}\{f(\lambda)\}(s) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} f(\lambda)\, e^{is\lambda}\, d\lambda .$$

Theorem 4.0.3 (Plancherel's Theorem)**.** *Consider $f \in L^1(\mathbb{R}^d)\bigcap L^2(\mathbb{R}^d)$. Then $\hat{f} \in L^2(\mathbb{R}^d)$ and*

$$\|\hat{f}\|_2 = (2\pi)^{d/2}\|f\|_2 \,.$$

Proof. See [20]. □

Since $L^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d)$ is dense in $L^2(\mathbb{R}^d)$, it is possible to extend the definition of Fourier transform to $f \in L^2(\mathbb{R}^d)$. Moreover, the Fourier transform is an automorphism in $L^2(\mathbb{R}^d)$ (see [20], [34], [35]).

Proposition 4.0.4 (Parseval Identity)**.** *Consider $f,g \in L^2(\mathbb{R}^d)$. Then*

$$\int_{\mathbb{R}^d} f(s)g(s)\,ds = (2\pi)^{-d}\int_{\mathbb{R}^d} \hat{f}(\lambda)\hat{g}(\lambda)\,d\lambda \,.$$

For a proof, see [20].

We should observe that the Fourier transform can be extended to generalized functions. Consider $f,g \in L^2(\mathbb{R}^d)$, then

$$(f,\hat{g}) = \int_{\mathbb{R}^d} f\hat{g}\,d\lambda = \int_{\mathbb{R}^d} \hat{\check{f}}\,\hat{g}\,d\lambda = (2\pi)^d\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \hat{\check{f}}\,\hat{g}\,d\lambda = (2\pi)^d\int_{\mathbb{R}^d} \check{f}g\,ds = (2\pi)^d d(\check{f},g)\,,$$

where we use the Parseval Identity. So, following [17], we generalize the Fourier transform for generalized functions $T \in \mathcal{S}^*$ in the following way

$$(\hat{T},f) := (2\pi)^d(T,\check{f}), \quad \text{for all } f \in L^2(\mathbb{R}^d).$$

In a similar way, we define the inverse Fourier transform as

$$(\check{T},f) = (2\pi)^{-d}(T,\hat{f}).$$

For instance, the Fourier transform of the Dirac delta function δ, is given by

$$(\hat{\delta},f) := (2\pi)^d(\delta,\check{f},) = \check{f}(0) = \int_{\mathbb{R}^d} f(s)\,ds = (1,f)$$

for all $f \in L^2(\mathbb{R}^d)$. We conclude that $\hat{\delta}(\lambda) = 1$.

The Fourier transform for a generalized random function ξ follows the same definition as for generalized functions,

$$(\hat{\xi},f) = (2\pi)^d(\xi,\check{f}) \qquad \text{and} \qquad (\check{\xi},f) = (2\pi)^{-d}(\xi,\hat{f})\,.$$

4.1 Solving a stochastic partial differential equation

Consider the linear SPDE

$$(k^2-\Delta)^{\alpha/2}\tau x = \mathcal{W} \tag{4.1.1}$$

where $s \in \mathbb{R}^d$, $\alpha = v + d/2$, $k > 0$, $\tau > 0$, $v > 0$, and $\mathcal{W}$ is a Gaussian white noise. To solve equation (4.1.1), we define the Fourier transform of a fractional Laplacian. In order to do that, consider the following,

$$\begin{aligned}\mathcal{F}\{(k^2-\Delta)\tau x\}(\lambda) &= \tau k^2\mathcal{F}\{x\}(\lambda) - \tau((i\lambda_1)^2 + \cdots + (i\lambda_d)^2)\mathcal{F}\{x\}(\lambda)\\ &= \tau(k^2 + \|\lambda\|^2)\mathcal{F}\{x\}(\lambda)\end{aligned} \tag{4.1.2}$$

where we use Proposition 4.0.4. We use (4.1.2) to define the Fourier transform of the fractional differential operator,

$$\mathcal{F}\{(k^2-\Delta)^{\alpha/2}\tau x\}(\lambda) = \tau(k^2+\|\lambda\|^2)^{\alpha/2}\mathcal{F}\{x\}(\lambda)\,. \tag{4.1.3}$$

Applying (4.1.3) to the SPDE (4.1.1), we obtain

$$\begin{aligned}&\mathcal{F}\{(k^2-\Delta)^{\alpha/2}\tau x\} = \mathcal{F}\{\mathcal{W}\}\\ \Leftrightarrow\ &\tau(k^2+\|\lambda\|^2)^{\alpha/2}\hat{x}(\lambda) = \hat{\mathcal{W}}\\ &\Leftrightarrow x(s) = \mathcal{F}^{-1}\left\{\frac{1}{\tau(k^2+\|\lambda\|^2)^{\alpha/2}}\hat{\mathcal{W}}\right\}(s)\,.\end{aligned} \tag{4.1.4}$$

If we define $G = \mathcal{F}^{-1}\left\{\frac{1}{\tau(k^2+\|\lambda\|^2)^{\alpha/2}}\right\}$, the solution of equation (4.1.1) can be written as a convolution,

$$x(s) = (\mathcal{W} * G)(s)\,. \tag{4.1.5}$$

Notice the following,

$$\int_{\mathbb{R}^d}\left(\frac{1}{\tau(k^2+\|\lambda\|^2)^{\alpha/2}}\right)^2 d\lambda = \int_{\mathbb{R}^d}\frac{1}{\tau^2(k^2+\|\lambda\|^2)^{\alpha}}\,d\lambda < \infty$$

for all $\alpha > 1$. We are interested in dimension $d \geq 2$, so in our setting $\alpha > 1$. We conclude that

$$\frac{1}{\tau(k^2+\|\lambda\|^2)^{\alpha/2}} \in L^2(\mathbb{R}^d)$$

therefore $G \in L^2(\mathbb{R}^d)$ as well as $G(s-\cdot) := G \circ h_s(\cdot)$ for every $s \in \mathbb{R}$, where $h_s(s') = (s-s')$ is a translation. For each $s \in \mathbb{R}^d$,

$$x(s) = (\mathcal{W}, G(s-\cdot))$$

and by the Definition 2.1.15 of the white noise, $(\mathcal{W}, G(s-.))$ is a Gaussian variable for all $s \in \mathbb{R}$, with zero mean and variance given by $\|G(s-.)\|_2^2 = \|G\|_2^2$.

4.2 Covariance function of $x(s)$

A very important aspect of the solution $x(s)$ is its covariance function. Consider $f, g \in L^2(\mathbb{R}^d)$. From (4.1.5),

$$\begin{aligned}
\mathbb{E}[(x,f)(x,g)] &= \mathbb{E}[(\mathcal{W}*G,f)(\mathcal{W}*G,g)] = \mathbb{E}[(\mathcal{W},f*G_{\circ})(\mathcal{W},g*G_{\circ})] \\
&= \mathbb{E}[(\mathcal{W},f*G)(\mathcal{W},g*G)] = (\delta,(f*G)*(g*G)_{\circ}) \\
&= \frac{(2\pi)^d}{(2\pi)^d}\left(\delta,\mathcal{F}^{-1}\{\mathcal{F}(f*G)\cdot\mathcal{F}(g_{\circ}*G)\}\right) = \frac{1}{(2\pi)^d}\left(\hat{\delta},\mathcal{F}(f*G)\cdot\mathcal{F}(g_{\circ}*G)\right) \\
&= \frac{1}{(2\pi)^d}\left(1,\hat{f}\,\hat{g}_{\circ}\,\frac{1}{\tau^2(k^2+\|\lambda\|^2)^{\alpha}}\right) = \frac{1}{(2\pi)^d}\left(\frac{1}{\tau^2(k^2+\|\lambda\|^2)^{\alpha}},\hat{f}\,\hat{g}_{\circ}\right) \\
&= \frac{1}{(2\pi)^d}\left(\frac{1}{\tau^2(k^2+\|\lambda\|^2)^{\alpha}},\mathcal{F}\{f*g_{\circ}\}\right) = \left(\mathcal{F}^{-1}\left\{\frac{1}{\tau^2(k^2+\|\lambda\|^2)^{\alpha}}\right\},f*g_{\circ}\right)
\end{aligned}$$

Let

$$H(s) = \mathcal{F}^{-1}\left\{\frac{1}{\tau^2(k^2+\|\lambda\|^2)^{\alpha}}\right\}(s).$$

We have the following,

$$\begin{aligned}
\mathbb{E}[(x,f)(x,g)] &= (H,f*g_{\circ}) = \int_{\mathbb{R}^d} H(s)\,f*g_{\circ}(s)\,ds \\
&= \int_{\mathbb{R}^d} H(s)\left(\int_{\mathbb{R}^d} f(s')g_{\circ}(s-s')\,ds'\right)ds = \iint_{\mathbb{R}^{2d}} H(s)f(s')g(s'-s)\,ds'\,ds \\
&= \iint_{\mathbb{R}^{2d}} H(s'-s)f(s')g(s)\,ds\,ds' = \iint_{\mathbb{R}^{2d}} H(s-s')f(s)g(s')\,ds\,ds'
\end{aligned}$$

therefore, Cov(x(s),x(s')) = $H(s-s')$. Following [38] and [39] the covariance function of the Gaussian field $x(s)$ is given by

$$\begin{aligned}
\mathrm{Cov}(x(s),x(s')) &= \mathcal{F}^{-1}\left\{\frac{1}{\tau^2(k^2+\|\lambda\|^2)^{\alpha}}\right\}(s-s') \\
&= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\frac{e^{i\lambda(s-s')}}{\tau^2(k^2+\|\lambda\|^2)^{\alpha}}\,d\lambda \\
&= \frac{\|s-s'\|^{v}\,K_v(k\|s-s'\|)}{2^{v-1+d/2}\,k^{v}\,\Gamma(v+d/2)\tau^2}, \qquad (4.2.1)
\end{aligned}$$

where K_v is the modified Bessel function of the second kind and order v, and Γ is the Gamma function.

We have the following,

$$H(s) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\frac{e^{i\lambda s}}{\tau^2(k^2+\|\lambda\|^2)^{\alpha}}\,d\lambda\,.$$

The function H(s) is positive definite. We call *spectral density function* of $H(s)$ to the function

$$f(\lambda) = \frac{1}{(2\pi)^d \tau^2 (k^2 + \|\lambda\|^2)^\alpha}.$$

This function is also called the *wave number spectrum* of the stationary solution $x(s)$.

Rearranging (4.2.1) and following [29], we obtain the *Matérn covariance*

$$\mathrm{Cov}(x(s), x(s')) = \frac{\sigma^2}{2^{v-1}\Gamma(v)} (k\|s-s'\|)^v K_v(k\|s-s'\|),$$

where

$$\sigma^2 = \frac{\Gamma(v)}{\Gamma\left(v + \frac{d}{2}\right)(4\pi)^{d/2} k^{2v} \tau^2}$$

is the marginal variance of the underlying process $x(s)$. We conclude that the solution $x(s)$ of the SPDE (4.1.1) is a GF with zero mean and Matérn covariance, also called *Matérn field*.

Gaussian fields with Matérn covariance are very important in spatial statistics because they describe a very large range of applications (see [5], [36]).

The parameter v is linked to the smoothness of the solution $x(s)$, The higher the value, the greater the smoothness (see [28]). The parameter k is a scaling parameter that determines the spatial correlation range. The empirically derived formula

$$\rho = \frac{\sqrt{8v}}{k},$$

called *pratical range*, defines ρ as the distance at which the correlation is around 0.1. For fixed v, the higher the k, the lower the distance ρ (see [28]).

4.3 Gaussian Markov random fields

In this subsection we follow [36], for the definition of Gaussian Markov random field (GMRF). First, we need to define *labelled graph*.

Definition 4.3.1. *A* labelled graph *is a tuple* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$*, where* $\mathcal{V} = \{1, ..., m\}$ *is an index label of m nodes* $\{s_i \in \mathbb{R}^d : 1, ..., m\} \subset D$*, and* $\mathcal{E}$ *is a set of edges* $\{i, j\}$*, with* $i \neq j$*, such that* $\{i, j\} \in \mathcal{E}$ *if and only if nodes* s_i *and* s_j *are connected by an edge.*

Definition 4.3.2. *A random vector* $\mathbf{x} = (x_1, \ldots, x_m)$ *is called a* Gaussian Markov random field *with respect to a labelled graph* $\mathcal{G}$*, with mean* (μ) *and precision matrix* $\mathbf{Q} > 0$*, if and only if its density has the form*

$$\pi(\mathbf{x}) = (2\pi)^{m/2} |\mathbf{Q}|^{1/2} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{Q} (\mathbf{x}-\boldsymbol{\mu})\right) \tag{4.3.1}$$

and

$$\mathbf{Q}_{ij} = 0 \quad \Leftrightarrow \quad \{i, j\} \notin \mathcal{E}.$$

for $i \neq j$*.*

Observe the following, given a Gaussian random vector $\mathbf{x} = (x_1, \ldots, x_m)$ with density given by (4.3.1), we can define a labelled graph $\tilde{\mathcal{G}} = (\mathcal{V}, \mathcal{E})$ such that $\tilde{\mathcal{V}} = \{1, \ldots, m\}$, and $Q_{ij} = 0$ if and only if $\{i, j\} \notin \tilde{\mathcal{E}}$. We conclude that any Gaussian random vector $\mathbf{x}$ is a GMRF (with respect to some graph $\tilde{\mathcal{G}}$).

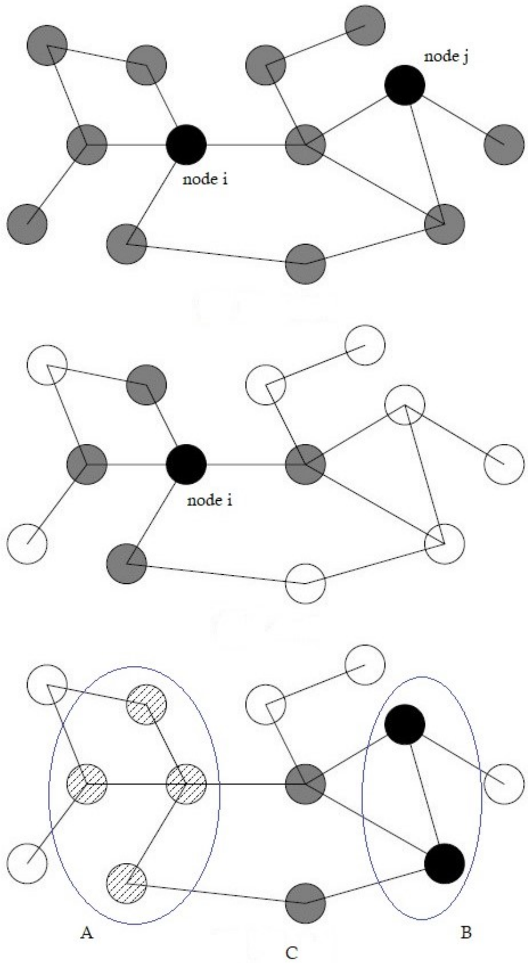


Figure 4.1: Top: pairwise Markov property. Middle: local Markov property. Bottom: global Markov property. See [36].

Figure 4.1 illustrates pairwise, local, and global Markov properties. If the nodes i and j are conditional independent given all the other nodes, we say that we have pairwise Markov property. If node i and the set of white nodes, are conditional independent given the neighbouring nodes $\mathcal{N}(i)$, in grey, we say that we have local Markov property. Finally, if set A (stripes) and set B (black) are conditional independent given the common neighbours, C (grey nodes), we say that we have global Markov property.

Theorem 4.3.3. *Let $\mathbf{x} = (x_1, \ldots, x_m)$ be a Gaussian random vector with mean $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q}$. Then, for $i \neq j$,*

$$x_i \perp x_j \mid \mathbf{x}_{-ij} \quad \Leftrightarrow \quad \mathbf{Q}_{ij} = 0.$$

where $\mathbf{x}_{-ij}$ stands for the vector obtained from $\mathbf{x}$, removing entries i and j.

We can find the proof of Theorem 4.3.3 in [36]. The proof relies on the following Theorem 4.3.4, which is not proved in [36]. Because of the lack of that proof, we present one here.

Theorem 4.3.4. *Consider variables x, y and z, with $\pi(z) > 0$. We have the following,*

$$x \perp y \mid z \quad \Leftrightarrow \quad \pi(x,y,z) = f(x,z)g(y,z)$$

for some functions f and g.

Proof. ($\Rightarrow$) We have the following,

$$\begin{aligned} x \perp y \mid z \quad \Leftrightarrow \quad \pi(x,y|z) &= \pi(x|z)\pi(y|z) \\ \frac{\pi(x,y,z)}{\pi(z)} &= \pi(x|z)\pi(y|z) \\ \pi(x,y,z) &= \pi(x|z)\pi(y|z)\pi(z). \end{aligned}$$

Consider $f(x,z) = \pi(x|z)$ and $g(y,z) = \pi(y|z)\pi(z)$. Therefore, $\pi(x,y,z) = f(x,z)g(y,z)$.

($\Leftarrow$) We have the following,

$$\pi(x|z) = \int \frac{\pi(x,y,z)}{\pi(z)}\,dy = \frac{f(x,z)}{\pi(z)} \int g(y,z)\,dy$$

and

$$\pi(y|z) = \int \frac{\pi(x,y,z)}{\pi(z)}\,dx = \frac{g(y,z)}{\pi(z)} \int f(x,z)\,dx$$

so we obtain

$$\pi(x|z)\pi(y|z) = \frac{f(x,z)g(y,z)}{\pi(z)^2} \iint f(x,z)g(y,z)\,dx\,dy = \frac{f(x,z)g(y,z)}{\pi(z)^2}\pi(z) = \pi(x,y|z).$$

Therefore, x and y are conditionally independent given z. □

Theorem 4.3.3 states that the precision matrix $\mathbf{Q}$ of a Gaussian random vector $\mathbf{x}$ indicates whether x_i and x_j are conditionally independent. Moreover, the zero entries of $\mathbf{Q}$ determine $\mathcal{G}$ and vice-versa.

To understand the Markov properties of a GMRF, let us define *neighbouring nodes*. Given a labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, define the neighbouring nodes of i as

$$\mathcal{N}(i) = \{l \in \mathcal{V} : \{i,l\} \in \mathcal{E}\}.$$

Given a set of nodes $A \subset \mathcal{V}$, we define

$$\mathcal{N}(A) = \left(\bigcup_{i \in A} \mathcal{N}(i) \right) \setminus A\,.$$

Consider $A, B, C \subset \mathcal{V}$ subsets of nodes. We say that C *separates* A and B if and only if A and B are disjoint non-empty sets such that

$$\mathcal{N}(A) \cap B = A \cap \mathcal{N}(B) = \varnothing \tag{4.3.2}$$

and C is given by

$$C = \mathcal{N}(A) \cap \mathcal{N}(B) .$$

This means the following, from (4.3.2) there is no edge between the nodes of A and the nodes of B, and C is the set of common neighbours to both A and B.

Theorem 4.3.5. *Let* $\mathbf{x}$ *a GMRF with respect to a labelled graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$*. Then the following Markov properties are equivalent.*

1. Pairwise Markov property*: if* $i \neq j$ *such that* $\{i, j\} \notin \mathcal{E}$*, then* $x_i \perp x_j \mid \mathbf{x}_{-ij}$.

2. Local Markov property*: for every* $i \in \mathcal{E}$ *we have* $x_i \perp \mathbf{x}_{-\{i, \mathcal{N}(i)\}} \mid \mathbf{x}_{\mathcal{N}(i)}$.

3. Global Markov property*: for every* $A, B, C \subset \mathcal{V}$ *such that* C *separates* A *and* B*, we have* $\mathbf{x}_A \perp \mathbf{x}_B \mid \mathbf{x}_C$.

We can find in [36] the illustration of the Markov properties, presented in Figure 1.

4.4 Finite Element Method

The Finite Element Method is designed to solve numerically differential equations, and obtain an approximation of the solution. Consider a domain $D \subset \mathbb{R}^d$. We are interested in a finite representation $\tilde{x}$ of the solution of SPDE (4.1.1), of the form

$$\tilde{x}(s) = \sum_{i=1}^{m} x_i \varphi_i(s)$$

where $\mathbf{x} = (x_1, \cdots, x_m)$ is a Gaussian random vector of weights, and $\{\varphi_i : i = 1, \ldots m\}$ is a set of finite element basis functions, defined by a chosen triangulation of the domain D, that is, a mesh. We present an example of a mesh in Figure 2a.

We call *mesh* to a labelled graph that divides the domain in (small) triangles. The basic idea of mesh is a set of nodes $\{s_i : i = 1, \ldots, m\} \subset D$, connected by edges, such that the domain D is divided in triangles.

Given a mesh $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of the domain $D \subset \mathbb{R}^d$, the Finite Element Method (FEM) defines a basis of finite element functions $\{\varphi_1(s), \ldots, \varphi_m(s)\}$ in the following way: for each mesh node n_i, the base element φ_i is a piece-wise linear function with $\varphi_i(n_i) = 1$, linearly decreasing to zero in each triangle with a vertex in n_i, as shown in Figure 2b.

In this work we consider the case where $\alpha = 2$. SPDE (4.1.1) will be solved in the weak sense, that is

$$((k^2 - \Delta)\, \tau x, g) = (\mathcal{W}, g), \tag{4.4.1}$$

for all $g \in \mathcal{S}$. The left hand side can be written as

$$((k^2 - \Delta)\, \tau x, g) = k^2 \tau \int_D x\, g\, ds - \tau \int_D \Delta x\, g\, ds . \tag{4.4.2}$$

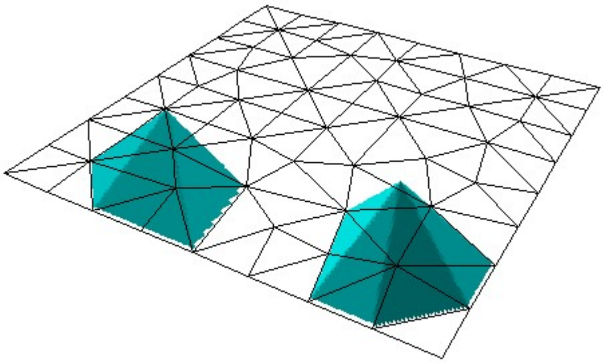


Figure 4.2: Two base element function. See [28].

Theorem 4.4.1 (Green's first identity)**.**

$$\int_D \Delta x\, g\, ds + \int_D \nabla x \nabla g\, ds = \int_{\partial D} g(\nabla x \cdot \mathbf{n})\, ds$$

For a proof, see [20]. Applying Green's Theorem 4.4.1 to (4.4.2), we obtain

$$((k^2 - \Delta)\,\tau x, g) = k^2\tau \int_D x\, g\, ds + \tau \int_D \nabla x \nabla g\, ds - \tau \int_{\partial\mathscr{D}} g(\nabla x \cdot \mathbf{n})\, ds \tag{4.4.3}$$

Also, SPDE (4.1.1) will be solved on a limited domain $\mathscr{D}$, so a proper boundary condition will be needed. Usually it is considered the Neumann condition, which states that

$$\frac{\partial x(s)}{\partial \mathbf{n}}|_{\partial D} = \nabla x \cdot \mathbf{n} = 0 \tag{4.4.4}$$

where $\mathbf{n}$ is the normal vector on the boundary ∂D. Therefore, equation (4.4.1) can be written in the following way,

$$k^2\tau \int_D x\, g\, ds + \tau \int_D \nabla x \nabla g\, ds = (\mathcal{W}, g) \tag{4.4.5}$$

Now consider the finite representation

$$\tilde{x}(s) = \sum_{i=1}^{m} x_i \varphi_i(s) \tag{4.4.6}$$

where $\mathbf{x} = (x_1, \cdots, x_m)$ is a Gaussian random vector of weights, to be determined. Notice that $\tilde{x}$ is an element of the Hilbert space H generated by the basis functions $\{\varphi_1, \ldots, \varphi_m\}$, for it is a linear combination of the basis-functions φ_i. Consider φ_j, and let

$$b_j := (\mathcal{W}, \varphi_j). \tag{4.4.7}$$

By Definition 2.1.15 of white noise, b_j is a zero-mean Gaussian random variable such that

$$\mathrm{Cov}(b_i, b_j) = \mathbb{E}[(\mathcal{W}, \varphi_i)(\mathcal{W}, \varphi_j)] = (\varphi_i, \varphi_j)\,. \tag{4.4.8}$$

Applying SPDE (4.1.1) to $\tilde{x}(s)$ in the weak sense, we obtain

$$((k^2-\Delta)\,\tau\tilde{x},\varphi_j) = (\mathcal{W},\varphi_j), \qquad \text{for all } j=1,\dots,n. \tag{4.4.9}$$

From (4.4.5) and (4.4.9) we obtain

$$\sum_{i=1}^{m} x_i k^2 \tau \int_D \varphi_i(s)\varphi_j(s)ds + \sum_{i=1}^{m} x_i \tau \int_D \nabla\varphi_i(s)\nabla\varphi_j(s)ds = b_j \qquad \text{for all } j=1,\dots,m. \tag{4.4.10}$$

We define the matrix $\mathbf{K}$ as

$$\mathbf{K}_{ij} = k^2 \int_D \nabla\varphi_i(s)\nabla\varphi_j(s)\,ds + \int_D \varphi_i(s)\varphi_j(s)\,ds.$$

and also we define the following matrices $\mathbf{C}$ and $\mathbf{G}$, as

$$\mathbf{C}_{ij} = \int_{\mathcal{D}} \varphi_i(s)\varphi_j(s)\,ds \quad \text{and} \quad \mathbf{G}_{ij} = \int_{\mathcal{D}} \nabla\varphi_i(s)\nabla\varphi_j(s)\,ds,$$

so $\boldsymbol{K} = k^2\mathbf{C}+\mathbf{G}$. Notice that $\mathbf{C}$, $\mathbf{G}$, and $\boldsymbol{K}$ are symmetric matrices. In this way, we can rewrite (4.4.10) in matricial form,

$$\tau(k^2\mathbf{C}+\mathbf{G})\mathbf{x} = \mathbf{b} \tag{4.4.11}$$

where $\mathbf{x} = (x_i,\dots,x_m)$ is the random vector of weights in (4.4.6), and $\mathbf{b} = (b_1,\dots,b_m)$ is the random vector with entries defined in (4.4.7). We can now calculate the precision matrix for the Gaussian random vector $\mathbf{x}$. From (4.4.8) and (4.4.11),

$$\tau(k^2\mathbf{C}+\mathbf{G})\mathbf{x} = \mathbf{b} \quad \Rightarrow \quad \text{Cov}(\tau(k^2\mathbf{C}+\mathbf{G})\mathbf{x}) = \mathbf{C}$$

so

$$\tau(k^2\mathbf{C}+\mathbf{G})\text{Cov}(\mathbf{x})\tau(k^2\mathbf{C}+\mathbf{G}) = \mathbf{C}$$
$$\Leftrightarrow \tau^{-1}(k^2\mathbf{C}+\mathbf{G})^{-1}\mathbf{Q}\tau^{-1}(k^2\mathbf{C}+\mathbf{G})^{-1} = \mathbf{C}^{-1}$$

where $\mathbf{Q} = \text{Cov}^{-1}(\mathbf{x})$ is the precision matrix of $\mathbf{x}$. Therefore

$$\begin{aligned}\mathbf{Q} &= \tau(k^2\mathbf{C}+\mathbf{G})\mathbf{C}^{-1}\tau(k^2\mathbf{C}+\mathbf{G})\\ &= \tau^2\mathbf{K}\mathbf{C}^{-1}\mathbf{K}\\ &= \tau^2(k^4\mathbf{C}+2k^2\mathbf{G}+\mathbf{G}\mathbf{C}^{-1}\mathbf{G}).\end{aligned}$$

In order to obtain computational gains, we substitute matrix $\mathbf{C}$ by the diagonal matrix $\tilde{\mathbf{C}}$, defined by

$$\tilde{\mathbf{C}}_{ii} = \int_D \varphi_i(s)\,ds = (\varphi_i(s),1).$$

For error estimates, see [29]). Then, the precision matrix will be given by

$$\tilde{\mathbf{Q}} = \tau^2\tilde{\mathbf{K}}\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{K}}$$

where

$$\tilde{\mathbf{K}} = k^2\tilde{\mathbf{C}} + \mathbf{G}\,.$$

Consider $i \neq j$. By construction of the basis element functions,

$$\{i,j\} \notin \mathscr{E} \Rightarrow G_{ij} = 0\,. \tag{4.4.12}$$

Moreover, $\tilde{\mathbf{C}}$ is a diagonal matrix, so, for $i \neq j$, we have $\tilde{\mathbf{K}}_{ij} = \mathbf{G}_{ij}$, and

$$\{i,j\} \notin \mathscr{E} \Rightarrow \tilde{\mathbf{K}}_{ij} = 0\,.$$

The reason for (4.4.12) is the following: first observe that $\operatorname{supp}\varphi_i$ is a finite union of triangle of the mesh, and $\operatorname{supp}\varphi_i \cap \operatorname{supp}\varphi_i \neq \varnothing$ if and only if nodes s_i and s_j share an edge of a triangle, which is equivalent to $\{i,j\} \in \mathscr{E}$ (observe Figure 2a, for instance nodes 26 and 39). However, in the case where $\{i,j\} \in \mathscr{E}$, we have the following,

$$\operatorname{supp}\varphi_i \cap \operatorname{supp}\varphi_j = \bigcup_{\gamma} T_\gamma$$

is also a union of triangles of the mesh (actually, two at most, because a mesh is a triangulation of the domain, but for now we only need to realize that it is a finite number of triangles). Then,

$$\mathbf{G}_{ij} = \int_{\mathscr{D}} \nabla\varphi_i(s)\nabla\varphi_j(s)\,ds\,ds = \sum_{\gamma}\int_{T_\gamma} \nabla\varphi_i(s)\nabla\varphi_j(s)\,ds\,.$$

The basis elements are (non-constant) piecewise linear functions in each triangle $T_\gamma \subset \operatorname{supp}\varphi_i \cap \operatorname{supp}\varphi_j$. Depending on the geometry of the mesh, we may have $\nabla\varphi_i(s)\nabla\varphi_j(s) < 0$, as well as $\nabla\varphi_i(s)\nabla\varphi_j(s) > 0$, so $\mathbf{G}_{ij}$ might be zero for neighbouring nodes i and j, such that $i \neq j$.

Let $k_{ij} := \tilde{\mathbf{K}}_{ij}$ and $c_i := \operatorname{diag}(\tilde{\mathbf{C}})_i > 0$. Then

$$\begin{aligned}\tilde{\mathbf{Q}}_{ij} &= \left(\tilde{\mathbf{K}}\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{K}}\right)_{ij} \\ &= (k_{i1},\ldots,k_{il},\ldots,k_{im})\cdot(c_1^{-1}k_{1j},\ldots,c_l^{-1}k_{lj},\ldots,c_j^{-1}k_{mj})\,.\end{aligned} \tag{4.4.13}$$

Again, for $i \neq l$, if $\{i,l\} \notin \mathscr{E}$, then $k_{il} = 0$. Therefore,

$$\mathcal{N}(i) \cap \{j, \mathcal{N}(j)\} = \varnothing \quad \Rightarrow \quad k_{il}\,k_{jl} = 0$$

for all $l \in \{1,\ldots m\}$. So,

$$\mathcal{N}(i) \cap \{j, \mathcal{N}(j)\} = \varnothing \quad \Rightarrow \quad \tilde{\mathbf{Q}}_{ij} = 0\,. \tag{4.4.14}$$

According to [5], this is a formulation of a GMRF, where we (re)define the neighbouring nodes as

$$\tilde{\mathcal{N}}(i) = \{j \in \mathscr{V} : \mathcal{N}(i) \cap \{j, \mathcal{N}(j)\} \neq \varnothing\}\,.$$

Notice that $\tilde{\mathcal{N}}(i)$ will contain all previously defined neighbours $\mathcal{N}(i)$, called *first order neighbours*, but also nodes that are connected by an edge with at least one element of $\mathcal{N}(i)$, the *second order neighbours*. Consider the labelled graph $\tilde{\mathcal{G}} = (\{1,\ldots,m\},\tilde{\mathcal{E}})$, where the set of edges $\tilde{\mathcal{E}}$ is defined by the condition

$$\{i,j\} \in \tilde{\mathcal{E}} \quad \Leftrightarrow \quad j \in \tilde{\mathcal{N}}(i)$$

for all $i \neq j$. So, if $\{i,j\} \notin \mathcal{E}$, then $\mathcal{N}(i) \cap \{j, \mathcal{N}(j)\} = \varnothing$. Therefore, condition (4.4.14) reads

$$\{i,j\} \notin \tilde{\mathcal{E}} \quad \Rightarrow \quad x_i \perp x_j \mid \mathbf{x}_{-ij} \,,$$

and we conclude that the pairwise Markov property is verified by $\mathbf{x}$.

4.5 Matrices C and G

In this section we explicitly calculate the matrices $\mathbf{C}$ and $\mathbf{G}$, needed to obtain the precision matrix $\mathbf{Q}$ od the GMRF $\mathbf{x}$.

Consider a bounded domain $D \in \mathbb{R}^2$, where we define a a triangulation of the domain called mesh. To create a mesh, we use the `R-INLA` package available in `R`, specially developed to perform approximate Bayesian inference (INLA approach), proposed by Rue et al. ([37]).

It should be mentioned that the calculations presented in this section were used to implement an algorithm in `R` for the construction of the matrices $\mathbf{C}$, $\mathbf{G}$, and $\mathbf{Q}$, independently of `R-INLA`. The use of `R-INLA` was restricted to creation of the mesh, and as motivated by the possibility to define several parameters like thickness of the mesh, the maximum length of the sides of each triangle in the mesh, etc., and most important, to get a matrix with the information of neighbouring points. Meshes may easily have thousands of nodes, so it made sense to use the `R-INLA` instruction for the mesh.

We call node to each vertex in the mesh, and two nodes are said to be neighbours if and only if they define a side of a triangle in the mesh. As said before, `R-INLA` can construct several types of meshes through the instruction `inla.mesh.2d`, and we can get the matrix of neighbouring nodes by `V <- meshname$graph$vv`. If node i is neighbour of node j, $\mathtt{V}_{ij} = 1$, and it is zero otherwise. Given a matrix of neighbours `V`, we can set a routine to define a three column matrix, where each row has the indexes of the three vertices of each triangle of the mesh, with one column per vertex.

To set some notation, let's say the mesh defines a finite set of triangles $\{T_\alpha : \alpha \in \mathcal{I}\}$, where $\mathcal{I}$ is a set of indexes. The FEM defines a piece-wise linear element base function φ_i, for each mesh node p_i, with $\varphi_i(p_i) = 1$, linearly decreasing to zero in each triangle with a vertex in p_i, as shown in Figure 4.2. Define the support of φ_i as $D_i = \{T_\alpha : \varphi_{i|_{T_\alpha}} \neq 0\}$. We have that D_i is the union of a small set of non overlapping triangles, $D_i = \{T_\alpha : \alpha \in \mathcal{I}_i\}$, where $\mathcal{I}_i \subset \mathcal{I}$ is a subset of indexes.

The next step is to create a `data.frame` in `R`, such that for each base element φ_i, we have the information of all triangles $T_\alpha \in D_i$, the analytical expression of φ_i for each of those T_α, as well as $\nabla\varphi_{i|_{T_\alpha}}$, and the area of T_α. All these quantities are needed to calculate matrices $\mathbf{C}$ and $\mathbf{G}$, in (5.4.7) and (5.4.8), respectively.

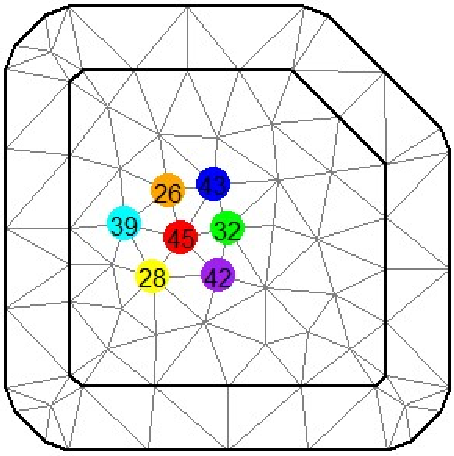


Figure 4.3: Node 45 and neighbours.

For instance, consider Figure 4.3, where we have node p_{45}, surrounded by its neighbouring nodes, p_{26}, p_{28}, p_{32}, p_{39}, p_{42}, and p_{43}. The domain D_{45} were $\varphi_{45} \neq 0$ is the union of six non overlapping triangles T_α, with $\alpha \in \{1,\dots,6\}$, defined by p_{45} and its neighbours. Therefore, we need to calculate $\mathbf{C}_{45,45}$, but also $\mathbf{G}_{45,j} = \mathbf{G}_{j,45}$ for $j \in \{26, 28, 32, 39, 42, 43, 45\}$.

Consider first of all, the problem of calculating $\mathbf{C}_{ii}$,

$$\mathbf{C}_{ii} = \iint_{D_i} \varphi_i(x,y)\, dxdy = \sum_{\alpha \in \mathcal{I}_i} \iint_{T_\alpha} \varphi_{i|_{T_\alpha}}(x,y)\, dxdy\,.$$

We need to compute the analytical expression of each restriction $\varphi_{i|_{T_\alpha}}$, which is a linear polynomial such that $\varphi_i(p_i) = 1$ and $\varphi_i(p_k) = \varphi_i(p_l) = 0$, where p_k and p_l are the remain vertices of T_α. Let $p_i = (a_i, b_i)$ be the $\mathbb{R}^2$ coordinates of node p_i. Then, the graphic of $\varphi_{i|_{T_\alpha}}$ is a triangle in $\mathbb{R}^3$, with vertices given by $(a_i, b_i, 1)$, $(a_k, b_k, 0)$, and $(a_l, b_l, 0)$. Given a normal vector to that graphic, for instance, $\mathrm{n}^i = (n_x^i, n_y^i, 1)$, we have the following,

$$n_x^i(a_i - x) + n_y^i(b_i - y) + (1 - z) = 0$$

for any point (x, y, z) belonging to the graphic of $\varphi_{i|_{T_\alpha}}$. Therefore,

$$\varphi_{i|_{T_\alpha}}(x,y) = n_x^i(a_i - x) + n_y^i(b_i - y) + 1\,. \tag{4.5.1}$$

So, we need to calculate the a normal vector n^i. Since, without loss of generality, we are assuming $n_z^i = 1$, we just need two equations to define n,

$$\begin{cases} \mathrm{n}^i \cdot (p_k - p_i) = 0 \\ \mathrm{n}^i \cdot (p_l - p_i) = 0 \end{cases} \Leftrightarrow \quad \begin{cases} n_x^i(a_k - a_i) + n_y^i(b_k - b_i) - 1 = 0 \\ n_x^i(a_l - a_i) + n_y^i(b_l - b_i) - 1 = 0 \end{cases}.$$

For simplicity, as we assume the third entry of the normal vector equals to 1, from now on $\mathrm{n}^i := (n_x^i, n_y^i)$. So, n^i is the solution of the following linear system

$$A_\alpha^i\, \mathrm{n}^i = \mathbf{1}$$

where **1** is a vector of 1's and

$$A_\alpha^i = \begin{bmatrix} a_k - a_i & b_k - b_i \\ a_l - a_i & b_l - b_i \end{bmatrix} .$$

We conclude that

$$\begin{bmatrix} n_x^i \\ n_y^i \end{bmatrix} = \frac{1}{\det(A_\alpha^i)} \begin{bmatrix} b_l - b_k \\ -a_l + a_k \end{bmatrix} \tag{4.5.2}$$

Notice that even though we swap p_k and p_l, the determinant in (4.5.2) will change sign, also does the vector it multiplies, so we get the same n^i, as we should. From (4.5.1) and (4.5.2),

$$\varphi_{i|_{T_\alpha}}(x,y) = \frac{1}{\det(A_\alpha^i)}((b_l - b_k)(a_i - x) + (-a_l + a_k)(b_i - y)) + 1 .$$

so

$$\nabla\varphi_{i|_{T_\alpha}}(x,y) = \frac{-1}{\det(A_\alpha^i)} \begin{bmatrix} b_l - b_k & -a_l + a_k \end{bmatrix} = \begin{bmatrix} -n_x^i & -n_y^i \end{bmatrix} = -(\mathrm{n}^i)^T$$

where the superscript T denotes the transpose. Consider now the triangle T defined by vertices $(0,0)$, $(1,0)$, and $(0,1)$, and let $f : T \to T_\alpha$ be a linear transformation such that

$$f(0,0) = p_i, \quad f(1,0) = p_k, \quad f(0,1) = p_l .$$

We have the following

$$(x,y) = f(u,v) = \begin{bmatrix} a_k - a_i & a_l - a_i \\ b_k - b_i & b_l - b_i \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} + \begin{bmatrix} a_i \\ b_i \end{bmatrix} . \tag{4.5.3}$$

Let B_α be the 2×2 matrix in (4.5.3). We have that $B_\alpha = (A_\alpha^i)^T$. Observe that

$$|\det(J_f(u,v))| = |\det(B_\alpha)| = |\det(J_{\tilde{f}}(u,v))|$$

where $\tilde{f} : T \to T_\alpha$ is any linear transformation such that

$$\tilde{f}(0,0) = f(p_{\sigma(i)}), \quad \tilde{f}(1,0) = f(p_{\sigma(k)}), \quad \tilde{f}(0,1) = f(p_{\sigma(l)})$$

where $\sigma : \{i,k,l\} \to \{i,k,l\}$ is any permutation of $\{i,k,l\}$. This is important to observe because it simplifies the calculations.

Applying change of variables in the integration, we have

$$\begin{aligned} \iint_{T_\alpha} \varphi_i(x,y)\, dxdy &= \iint_T \varphi_i(f(u,v))|\det(B_\alpha)|\, dudv \\ &= \iint_T (-u - v + 1)|\det(B_\alpha)|\, dudv \\ &= \frac{|\det(B_\alpha)|}{6} \end{aligned} \tag{4.5.4}$$

therefore,

$$\mathbf{C}_{ii} = \sum_{T_\alpha \in D_i} \iint_{T_\alpha} \varphi_i(x,y)\, dxdy = \sum_{T_\alpha \in D_i} \frac{|\det(B_\alpha)|}{6} . \tag{4.5.5}$$

Also, we have

$$\begin{aligned}\iint_{T_\alpha} \|\nabla\varphi_i(x,y)\|^2\, dxdy &= \iint_{T_\alpha} \|\mathrm{n}^i\|^2\, dxdy \\ &= \iint_{T} \|\mathrm{n}^i\|^2 |\det(B_\alpha)|\, dudv \\ &= \frac{\|\mathrm{n}^i\|^2 |\det(B_\alpha)|}{2}\end{aligned}$$

so

$$\mathbf{G}_{ii} = \sum_{T_\alpha \in D_i} \iint_{T_\alpha} \|\nabla\varphi_i(x,y)\|^2\, dxdy = \sum_{T_\alpha \in D_i} \frac{\|\mathrm{n}^i\|^2 |\det(B_\alpha)|}{2}\,. \tag{4.5.6}$$

Moreover, consider now two base element functions φ_i and φ_j such that $D_{ij} = D_i \cap D_j \neq \varnothing$. In this case, D_{ij} is the union of one or two non overlapping triangles. Consider $T_\alpha \subset D_{ij}$, defined by vertices p_i, p_j, and p_l. Observe that

$$\nabla\varphi_{i|_{T_\alpha}}(x,y) = \frac{-1}{\det(A^i_\alpha)}\begin{bmatrix} b_l - b_j & -a_l + a_j \end{bmatrix} = -(\mathrm{n}^i)^T$$

and

$$\nabla\varphi_{j|_{T_\alpha}}(x,y) = \frac{-1}{\det(A^j_\alpha)}\begin{bmatrix} b_l - b_i & -a_l + a_i \end{bmatrix} = -(\mathrm{n}^j)^T\,.$$

Therefore,

$$\begin{aligned}\iint_{T_\alpha} \nabla\varphi_i \cdot \nabla\varphi_j\, dxdy &= \iint_{T} \mathrm{n}^i \cdot \mathrm{n}^j |\det(B_\alpha)| dudv \\ &= \frac{|\det(B_\alpha)|}{2}\mathrm{n}^i \cdot \mathrm{n}^j\end{aligned}$$

Notice that, the therefore,

$$\mathbf{G}_{ij} = \sum_{T_\alpha \subset D_{ij}} \frac{|\det(B_\alpha)|}{2}\mathrm{n}^i \cdot \mathrm{n}^j\,. \tag{4.5.7}$$

5

Integrated nested Laplace approximation methodology applied to wind velocity data

The main goal of this chapter is to predict the wind velocity, using Bayesian inference. We present all theoretical results and calculations, not explicitly presented in the literature, that supports INLA methodology, central for doing approximate Bayesian inference, and apply that methodology to estimate a spatial model for our wind data set ([26]).The results are encouraging, and open new lines of investigation, such as applying statistical methods to study the solution of stochastic partial differential equations.

We start by defining a specific class of hierarchical models, called latent Gaussian models. An hierarchical model combines different levels of information, for instance the observations are assumed to follow some distribution which depends on some set of latent random variables $\boldsymbol{\mu}$, following a distribution depending on a set of hyperparameters $\boldsymbol{\theta}$. Hierarchical models are defined in Subsection 2.2.2. In this work, we are interested in the special case of latent Gaussian models with one spatial effect, defined in Section 5.1, which are formulated as follows,

$$
\begin{aligned}
y_i|\mu_i,\boldsymbol{\theta} &\sim \pi(y_i|\mu_i,\boldsymbol{\theta})\,, \quad i=1,\dots,n \\
g(\mu_i) = \eta_i &:= \alpha + \sum_{l=1}^{L} z_i^l \beta_l + x(s_i) \\
\mathbf{x}|\boldsymbol{\theta} &\sim GF(\mathbf{0},\mathbf{Q}^{-1}(\boldsymbol{\theta})) \\
\boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta})
\end{aligned}
$$

were the observations $\mathbf{y}=(y_1,...,y_n)$ of some spatial fenomena at locations $(s_1,\dots,s_n)$, are assumed to follow a distribution $\pi(y_i|\mu_i,\boldsymbol{\theta})$. The parameter μ_i, usually taken as $E[y_i]$, is linked to a so called *predictor* η_i, defined above. Moreover, the vector of all latent variables $\mathbf{x}=(\alpha,\beta_1,\dots,\beta_L,x(s_1),\dots,x(s_n))$, conditional to a set of parameters $\boldsymbol{\theta}$, is assumed to be a *Gaussian field* with $\mathbf{0}$ mean and precision matrix $\mathbf{Q}(\boldsymbol{\theta})$. Here, we consider one spatial effect, $x(s)$. Moreover, $\pi(\boldsymbol{\theta})$ is the prior distribution of the hyperparameters.

The main goal is to obtain the posterior distributions $\pi(\boldsymbol{\theta}|\mathbf{y})$ and $\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})$, first theoretically, then computationally for the examples in Sections 5.4 and 5.5, and to estimate the mean and standard deviation of the spatial effect, $x(s)$, and consequently, the mean $\mu_i = E[y_i]$.

5.1 Latent Gaussian model with one spatial effect

Consider a domain $D \subset \mathbb{R}^d$, and a vector $\mathbf{y} = (y_1,\ldots,y_n)$ of observed data at locations $\mathbf{s} = (s_1,\ldots,s_n)$, with $s_i \in D$, for all $i = 1,\ldots,n$. Consider a model with the following predictor, with one spatial random effect,

$$\eta_i = \alpha + \sum_{l=1}^{L} z_i^l \beta_l + x(s_i),$$

where the spatial dependence of the observations is modelled by the Gaussian field $x(s)$. In this case, the vector of latent variables is

$$\mathbf{x} = (\alpha, \beta_1, \ldots, \beta_L, x(s_1), \ldots, x(s_n)),$$

where $\dim(\mathbf{x}) = m$. We assign a Gaussian prior distribution to the vector of all latent variables, $\mathbf{x}$, therefore, this model is called a latent Gaussian model. The distribution $\pi(\mathbf{x}|\boldsymbol{\theta})$ of the latent variables $\mathbf{x}$, is assumed to be Gaussian with zero mean and precision matrix $\mathbf{Q}(\boldsymbol{\theta})$. The observations $\mathbf{y} = (y_1,\ldots,y_n)$ are assumed to be conditionally independent given $\mathbf{x}$ and the parameters $\boldsymbol{\theta}$, and belonging to the same distribution family. The vector of all parameters, $\boldsymbol{\theta}$, may not be Gaussian distributed. Hence, our model can be written in the following way,

$$\begin{aligned} y_i|\mu_i,\boldsymbol{\theta} &\sim \pi(y_i|\mu_i,\boldsymbol{\theta}), \quad i = 1,\ldots,n \\ g(\mu_i) = \eta_i &:= \alpha + \sum_{l=1}^{L} z_i^l \beta_l + x(s_i) \\ \mathbf{x}|\boldsymbol{\theta} &\sim GF(\mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta})) \\ \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}). \end{aligned} \tag{5.1.1}$$

The conditional distribution of the observations $\mathbf{y}$ is given by

$$\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}) = \prod_{i=1}^{n} \pi(y_i|\eta_i,\boldsymbol{\theta}) = \prod_{i=1}^{n} \pi(y_i|\mathbf{x}_i,\boldsymbol{\theta}), \tag{5.1.2}$$

where $\mathbf{x}_i = (\alpha, \boldsymbol{\beta}, x(s_i))$, and the conditional distribution of $\mathbf{x}$ is given by

$$\pi(\mathbf{x}|\boldsymbol{\theta}) = (2\pi)^{-m/2} |\mathbf{Q}(\boldsymbol{\theta})|^{1/2} \exp\left\{-\frac{1}{2}\mathbf{x}^T \mathbf{Q}(\boldsymbol{\theta})\mathbf{x}\right\} \tag{5.1.3}$$

where $\dim(\mathbf{x}) = m$. The number of parameters has to be small, say $\dim(\boldsymbol{\theta}) = q \le 6$, for computational efficiency (see [37]).

Applying Bayes' Theorem,

$$\begin{aligned}\pi(\mathbf{x},\boldsymbol{\theta}|\mathbf{y}) &= \frac{\pi(\mathbf{x},\boldsymbol{\theta},\mathbf{y})}{\pi(\mathbf{y})}\\ &= \frac{\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\pi(\mathbf{x},\boldsymbol{\theta})}{\pi(\mathbf{y})}\\ &= \frac{\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathbf{y})}\\ &\propto \pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\end{aligned}$$

and considering (5.1.2) and (5.1.3), we obtain

$$\begin{aligned}\pi(\mathbf{x},\boldsymbol{\theta}|\mathbf{y}) &\propto \pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\prod_{i=1}^{n}\pi(y_i|\mathbf{x}_i,\boldsymbol{\theta})\\ &\propto \pi(\boldsymbol{\theta})|\mathbf{Q}(\boldsymbol{\theta})|^{1/2}\exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x}\right\}\prod_{i=1}^{n}\pi(y_i|\mathbf{x}_i,\boldsymbol{\theta})\\ &\propto \pi(\boldsymbol{\theta})|\mathbf{Q}(\boldsymbol{\theta})|^{1/2}\exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x}+\sum_{i=1}^{n}\log\{\pi(y_i|\mathbf{x}_i,\boldsymbol{\theta})\}\right\}. \qquad (5.1.4)\end{aligned}$$

Given the distribution $\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})$ and the priors $\pi(\mathbf{x}|\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta})$, we use Bayesian inference and Laplace approximations to obtain the posterior distributions $\pi(\mathbf{x}|\mathbf{y})$ and $\pi(\boldsymbol{\theta}|\mathbf{y})$, and their marginals $\pi(x_i|\mathbf{y})$ and $\pi(\theta_j|\mathbf{y})$, where x_i is the i^{th} entry of $\mathbf{x}$, for $i \in \{1,\dots,m\}$, and θ_j is the j^{th} entry of $\boldsymbol{\theta}$, for $j \in \{1,\dots,q\}$.

We can write the posterior distributions in the following way,

$$\begin{aligned}\pi(x_i|\mathbf{y}) &= \int \pi(x_i|\boldsymbol{\theta},\mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y})\,d\boldsymbol{\theta}\\ \pi(\theta_j|\mathbf{y}) &= \int \pi(\boldsymbol{\theta}|\mathbf{y})\,d\boldsymbol{\theta}_{-j}.\end{aligned}$$

5.2 The INLA methodology

In order to approximate $\pi(x_i|\mathbf{y})$ and $\pi(\theta_j|\mathbf{y})$ we will use approximations for the distributions $\pi(x_i|\boldsymbol{\theta},\mathbf{y})$ and $\pi(\boldsymbol{\theta}|\mathbf{y})$, and the integrals in (5.1) will be approximated using numerical integration. The INLA methodology calculates numerical approximations to the posterior distributions $\pi(\boldsymbol{\theta}|\mathbf{y})$ and $\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})$, based on the Gaussian approximation given by the Laplace method. The first step is to calculate an approximation for the distribution of the hyperparameters, $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$. We have the following

$$\begin{aligned}\pi(\boldsymbol{\theta}|\mathbf{y}) &= \frac{\pi(\mathbf{x},\boldsymbol{\theta}|\mathbf{y})}{\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})} \\ &= \frac{\pi(\mathbf{x},\boldsymbol{\theta}|\mathbf{y})\pi(\mathbf{y})}{\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})\pi(\mathbf{y})} \\ &= \frac{\pi(\mathbf{x},\boldsymbol{\theta},\mathbf{y})}{\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})\pi(\mathbf{y})} \\ &= \frac{\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})\pi(\mathbf{y})} \\ &\propto \frac{\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})} \\ &\approx \left.\frac{\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})}\right|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})} =: \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}). \end{aligned} \tag{5.2.1}$$

where $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})$ is the Gaussian approximation given by the Laplace method, explained in Section 5.3, and $\mathbf{x}^*(\boldsymbol{\theta})$ is the mode of that distribution, for a given $\boldsymbol{\theta}$. Observe that the left hand-side of (5.2.1) is a kernel of a certain density function, and $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ is by definition that density function.

The second step is to approximate $\pi(x_i|\boldsymbol{\theta},\mathbf{y})$. In this case, a possible approach would be to approximate it using the marginal distributions from $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})$, but the approximation in such a way is not a good one (see [37]). A second strategy would be to consider $\mathbf{x} = (x_i, \mathbf{x}_{-i})$, and use again the Laplace approximation to obtain

$$\tilde{\pi}(x_i|\boldsymbol{\theta},\mathbf{y}) \propto \left.\frac{\pi(\mathbf{x},\boldsymbol{\theta},\mathbf{y})}{\tilde{\pi}_G(\mathbf{x}_{-i}|x_i,\boldsymbol{\theta},\mathbf{y})}\right|_{\mathbf{x}_{-i}=\mathbf{x}^*_{-i}(x_i,\boldsymbol{\theta})} \tag{5.2.2}$$

where $\tilde{\pi}_G(\mathbf{x}_{-i}|x_i,\boldsymbol{\theta},\mathbf{y})$ is the Laplace Gaussian approximation to $\pi(\mathbf{x}_{-i}|x_i,\boldsymbol{\theta},\mathbf{y})$ and $\mathbf{x}^*_{-i}(x_i,\boldsymbol{\theta})$ is its modal configuration. According to [37], is computationally expensive to compute the Laplace Gaussian approximation $\tilde{\pi}_G(\mathbf{x}_{-i}|x_i,\boldsymbol{\theta},\mathbf{y})$, as it must be calculated for each x_i and $\boldsymbol{\theta}$, so the authors explore two modifications to (5.2.2) to reduce computational costs. A third option presented by the authors is a *simplified Laplace approximation* based on a Taylor's series expansion of $\tilde{\pi}(x_i|\boldsymbol{\theta},\mathbf{y})$ in equation (5.2.2), and includes a mixing term, for instance a cubic spline, to increase the fit. For further details, see [37].

5.3 Laplace method for Gaussian approximations

The Laplace method uses second order Taylor expansion to provide the Gaussian approximation $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})$ in equation (5.2.1). Consider the case with no fixed effects, where y_i is an observation with error of $x(s_i)$, $i = 1,\ldots,n$. In this case, $\mathbf{x} = (x_1,\ldots,x_n) = (x(s_1),\ldots,x(s_n))$. Notice that

$$\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y}) = \frac{\pi(\mathbf{x},\boldsymbol{\theta}|\mathbf{y})}{\pi(\boldsymbol{\theta}|\mathbf{y})} \propto \pi(\mathbf{x},\boldsymbol{\theta}|\mathbf{y})$$

and from (5.1.4),

$$\pi(\mathbf{x},\boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta})|\mathbf{Q}(\boldsymbol{\theta})|^{1/2} \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x} + \sum_{i=1}^{n} \log\{\pi(y_i|x_i,\boldsymbol{\theta})\}\right\},$$

therefore

$$\begin{aligned}\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y}) &\propto \pi(\boldsymbol{\theta})|\mathbf{Q}(\boldsymbol{\theta})|^{1/2}\exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x}+\sum_{i=1}^{n}\log\{\pi(y_i|x_i,\boldsymbol{\theta})\}\right\}\\ &\propto \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x}+\sum_{i=1}^{n}\log\{\pi(y_i|x_i,\boldsymbol{\theta})\}\right\}\\ &= \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x}+g(\mathbf{x})\right\}, \end{aligned} \tag{5.3.1}$$

where $g(\mathbf{x}) := \sum_{i=1}^{n} g_i(x_i)$, and $g_i(x_i) = \log\{\pi(y_i|x_i,\boldsymbol{\theta})\}$ is a function that does not depend of x_j for $j \neq i$. We have the following,

$$\nabla g(\mathbf{x}) = \left[\left.\frac{\partial \log\{\pi(y_1|x_1,\boldsymbol{\theta})\}}{\partial x_1}\right|_{x_1} \quad \cdots \quad \left.\frac{\partial \log\{\pi(y_n|x_n,\boldsymbol{\theta})\}}{\partial x_n}\right|_{x_n}\right]$$

and the Hessian matrix $\mathbf{H}$ of $g(\mathbf{x})$ is diagonal,

$$\mathbf{H}(\mathbf{x}) = \mathrm{diag}\left(\left.\frac{\partial^2 \log\{\pi(y_1|x_1,\boldsymbol{\theta})\}}{\partial x_1^2}\right|_{x_1}, \quad \cdots \quad, \left.\frac{\partial^2 \log\{\pi(y_n|x_n,\boldsymbol{\theta})\}}{\partial x_n^2}\right|_{x_n}\right).$$

Let $\mathbf{x}^* = \mathbf{x}^*(\boldsymbol{\theta})$ be the mode of $\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})$. The second order Taylor approximation for $g(\mathbf{x})$ around the mode is

$$g(\mathbf{x}) \approx g(\mathbf{x}^*) + \nabla g(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*) + \frac{1}{2}(\mathbf{x}-\mathbf{x}^*)^T\mathbf{H}(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*). \tag{5.3.2}$$

Considering (5.3.1) and (5.3.2),

$$\begin{aligned}\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y}) &\approx \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x}+g(\mathbf{x}^*)+\nabla g(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)+\frac{1}{2}(\mathbf{x}-\mathbf{x}^*)^T\mathbf{H}(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)\right\}\\ &= \exp\{g(\mathbf{x}^*)\}\exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x}+\nabla g(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)+\frac{1}{2}(\mathbf{x}-\mathbf{x}^*)^T\mathbf{H}(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)\right\}\\ &\propto \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x}+\nabla g(\mathbf{x}^*)\mathbf{x}+\frac{1}{2}(\mathbf{x}-\mathbf{x}^*)^T\mathbf{H}(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)\right\}. \end{aligned} \tag{5.3.3}$$

In one hand, we have,

$$\begin{aligned}(\mathbf{x}-\mathbf{x}^*)^T\mathbf{Q}(\mathbf{x}-\mathbf{x}^*) &= \mathbf{x}^T\mathbf{Q}(\mathbf{x}-\mathbf{x}^*) - (\mathbf{x}^*)^T\mathbf{Q}(\mathbf{x}-\mathbf{x}^*)\\ &= \mathbf{x}^T\mathbf{Q}\mathbf{x} - 2\,(\mathbf{x}^*)^T\mathbf{Q}\mathbf{x} + (\mathbf{x}^*)^T\mathbf{Q}\mathbf{x}^* \end{aligned} \tag{5.3.4}$$

where $\mathbf{Q} = \mathbf{Q}(\boldsymbol{\theta})$ for simplicity of notation. In the other hand, the mode is the maximum of the density probability function (for a differentiable density function), which implies[1]

$$\begin{aligned} &\nabla_{\mathbf{x}}\left(-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}+\nabla g(\mathbf{x}^*)\mathbf{x}+\frac{1}{2}(\mathbf{x}-\mathbf{x}^*)^T\mathbf{H}(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)\right)\Big|_{\mathbf{x}=\mathbf{x}^*} = \mathbf{0}\\ &\Leftrightarrow \left(-\mathbf{x}^T\mathbf{Q}+\nabla g(\mathbf{x}^*)+(\mathbf{x}-\mathbf{x}^*)^T\mathbf{H}(\mathbf{x}^*)\right)\Big|_{\mathbf{x}=\mathbf{x}^*} = \mathbf{0}\\ &\Leftrightarrow (\mathbf{x}^*)^T\mathbf{Q} = \nabla g(\mathbf{x}^*) \end{aligned} \tag{5.3.5}$$

[1] Alternatively, we could have calculated the gradient of (5.3.3).

From (5.3.4) and (5.3.5) we obtain

$$(\mathbf{x}-\mathbf{x}^*)^T\mathbf{Q}(\mathbf{x}-\mathbf{x}^*) = \mathbf{x}^T\mathbf{Q}\mathbf{x} - 2\nabla g(\mathbf{x}^*)\mathbf{x} + \nabla g(\mathbf{x}^*)\mathbf{x}^*, \tag{5.3.6}$$

therefore, from (5.3.3) and (5.3.6), and taking $\mathbf{D} = -\mathbf{H}$, we conclude that

$$\pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y}) \propto \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{x}^*)^T(\mathbf{Q}+\mathbf{D})(\mathbf{x}-\mathbf{x}^*)\right\}, \tag{5.3.7}$$

which is the kernel of a multivariate Normal distribution with mode $\mathbf{x}^* = \mathbf{x}^*(\boldsymbol{\theta})$ and covariance matrix $(\mathbf{Q}(\boldsymbol{\theta})+\mathbf{D})^{-1}$.

The iterative process to approximate the mode $\mathbf{x}^*(\boldsymbol{\theta})$ and the covariance $(\mathbf{Q}(\boldsymbol{\theta})+\mathbf{D})^{-1}$ is the following: consider an initial guess for the mode, $\mathbf{x}^{(0)} = (x_1^{(0)},\dots,x_n^{(0)})$, and approximate each $g_i(x_i)$ around $x_i^{(0)}$ using the second order Taylor approximation for each $g_i(x_i)$,

$$\begin{aligned} g_i(x_i) &\approx g(x_i^{(0)}) + (x_i - x_i^{(0)})\left.\frac{\partial g_i}{\partial x_i}\right|_{x_i=x_i^{(0)}} - \frac{1}{2}(x_i - x_i^{(0)})^2\left.\frac{\partial^2 g_i}{\partial x_i^2}\right|_{x_i=x_i^{(0)}} \\ &= a_i + b_i x_i - \frac{1}{2}c_i x_i^2 \end{aligned} \tag{5.3.8}$$

where

$$a_i = g(x_i^{(0)}) - x_i^{(0)}\left.\frac{\partial \log\{\pi(y_i|x_i,\boldsymbol{\theta})\}}{\partial x_i}\right|_{x_i=x_i^{(0)}} - \frac{1}{2}(x_i^{(0)})^2\left.\frac{\partial^2 \log\{\pi(y_i|x_i,\boldsymbol{\theta})\}}{\partial x_i^2}\right|_{x_i=x_i^{(0)}}$$

$$b_i = \left.\frac{\partial \log\{\pi(y_i|x_i,\boldsymbol{\theta})\}}{\partial x_i}\right|_{x_i=x_i^{(0)}} + x_i^{(0)}\left.\frac{\partial^2 \log\{\pi(y_i|x_i,\boldsymbol{\theta})\}}{\partial x_i^2}\right|_{x_i=x_i^{(0)}}$$

$$c_i = \left.\frac{\partial^2 \log\{\pi(y_i|x_i,\boldsymbol{\theta})\}}{\partial x_i^2}\right|_{x_i=x_i^{(0)}}$$

Let

$$\mathbf{a}^0 = \sum_{i=1}^n a_i, \quad \mathbf{b}^0 = (b_1,\dots,b_n), \quad \text{and} \quad \mathbf{C}^0 = \operatorname{diag}(c_1,\dots,c_n),$$

where $\operatorname{diag}(c_1,\dots,c_n)$ denotes the diagonal matrix with diagonal entries given by $(c_1,\dots,c_n)$. From (5.3.1),

$$\begin{aligned} \pi(\mathbf{x}|\boldsymbol{\theta},\mathbf{y}) &\propto \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x} + g(\mathbf{x})\right\} \\ &\approx \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x} + \mathbf{a}^0 + (\mathbf{b}^0)^T\mathbf{x} - \frac{1}{2}c_i x_i^2\right\} \\ &\propto \exp\left\{-\frac{1}{2}\mathbf{x}^T(\mathbf{Q}(\boldsymbol{\theta}) + \mathbf{C}^0)\mathbf{x} + (\mathbf{b}^0)^T\mathbf{x}\right\}. \end{aligned} \tag{5.3.9}$$

We obtain in equation (5.3.9) the kernel of a Gaussian density function with precision matrix $\mathbf{Q}(\boldsymbol{\theta}) + \mathbf{C}^0$, and mode $\mathbf{x}^{(1)}$ given by the solution of

$$(\mathbf{Q}(\boldsymbol{\theta}) + \mathbf{C}^0)\mathbf{x}^{(1)} = \mathbf{b}^0. \tag{5.3.10}$$

Next, we approximate each $g(x_i)$ around $x_i^{(1)}$, where $x_i^{(1)}$ is the i^{th} entry of $x^{(1)}$, and obtain a new kernel of a Gaussian distribution, with precision matrix $\mathbf{Q}(\boldsymbol{\theta})+\mathbf{C}^1$ and mode given by the solution of

$$(\mathbf{Q}(\boldsymbol{\theta}) + \mathbf{C}^1)\mathbf{x}^{(2)} = \mathbf{b}^1. \tag{5.3.11}$$

5.4 One-dimensional example

In Chapter 4 we have discussed that the solution of the SPDE

$$(k^2 - \Delta)^{\alpha/2} \tau x = \mathcal{W} \tag{5.4.1}$$

is a Gaussian field with Matérn covariance

$$\mathrm{Cov}(x(s), x(s')) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (k\|s - s'\|)^\nu K_\nu(k\|s - s'\|) \tag{5.4.2}$$

where

$$\begin{cases} \nu = \alpha - d/2 \\ \sigma^2 = \dfrac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2} k^{2\nu} \tau^2} \end{cases} . \tag{5.4.3}$$

We have that ν is a smoothness parameter, k is a scale parameter, and τ controls the variance. Moreover, $\mathcal{W}$ is a Gaussian white noise process, and σ^2 is the marginal variance of the process.

The default value for α in `R-INLA` is $\alpha = 2$, and for this example, $d = 1$, so we have

$$\begin{cases} \nu = \dfrac{3}{2} \\ \sigma^2 = \dfrac{1}{4k^3\tau^2} \end{cases} . \tag{5.4.4}$$

Suppose that we observe data $y_i = y(s_i)$ at location $s_i = -2 + 0.05(i - 1) \in [-2, 3]$, for $i \in \{1, \ldots, n\}$, with $n = 101$, that was generated by some underlying Gaussian field $x(s)$ that cannot be directly observed. Then,

$$y_i = x(s_i) + e_i \ ,$$

with e_i independent of e_j, for all $i \neq j$, identically distributed, and with zero mean. Assume that $x(s)$ is the solution of the SPDE (5.4.1), for some τ and k. Furthermore, assume that the solution $x(s)$ is approximated using the FEM through a basis function representation, defined on the partition $\{-2, -1, -0.5, 0, 1, 1.5, 2, 3\}$ of the domain. Each node of the partition defines a piece-wise linear function, with value equal to one at the nodes, linearly decreasing to zero until the neighbouring nodes, and zero elsewhere. Figure 5.1 is a representation of the basis functions, we can find a similar one in [28].

Then

$$x(s) \approx \sum_{i=1}^{8} x_i \varphi_i(s)$$

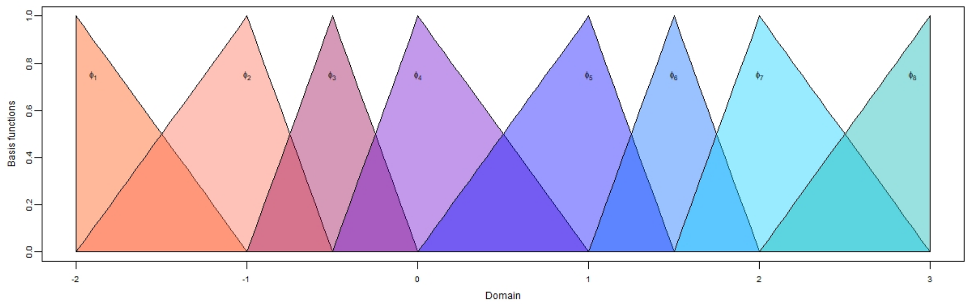


Figure 5.1: Basis Functions

where $\{\varphi_1, \ldots, \varphi_8\}$ is the set of basis functions and $\{x_1, \ldots, x_8\}$ are Gaussian distributed weights, with zero mean. The basis functions for our example are the following,

$$\varphi_1(x) = \begin{cases} -x-1 & , x \in [-2,-1] \\ 0 & , \text{otherwise.} \end{cases} \qquad \varphi_2(x) = \begin{cases} x+2 & , x \in [-2,-1[\\ -2x-1 & , x \in [-1,-0.5] \\ 0 & , \text{otherwise.} \end{cases}$$

$$\varphi_3(x) = \begin{cases} 2x+2 & , x \in [-1,-0.5[\\ -2x & , x \in [-0.5,0] \\ 0 & , \text{otherwise.} \end{cases} \qquad \varphi_4(x) = \begin{cases} 2x+1 & , x \in [-0.5,0[\\ -x+1 & , x \in [0,1] \\ 0 & , \text{otherwise.} \end{cases} \tag{5.4.5}$$

$$\varphi_5(x) = \begin{cases} x & , x \in [0,1[\\ -2x+3 & , x \in [1,1.5] \\ 0 & , \text{otherwise.} \end{cases} \qquad \varphi_6(x) = \begin{cases} 2x-2 & , x \in [1,1.5[\\ -2x+4 & , x \in [1.5,2] \\ 0 & , \text{otherwise.} \end{cases}$$

$$\varphi_7(x) = \begin{cases} 2x-3 & , x \in [1.5,2[\\ -x+3 & , x \in [2,3] \\ 0 & , \text{otherwise.} \end{cases} \qquad \varphi_8(x) = \begin{cases} x-2 & , x \in [2,3] \\ 0 & , \text{otherwise.} \end{cases}$$

Considering $\alpha = 2$ and using Neumann boundary conditions, the vector of weights $\mathbf{x} = (x_1, ..., x_7)$ has precision matrix given by

$$\mathbf{Q} = \tau^2(k^4\mathbf{C} + 2k^2\mathbf{G} + \mathbf{G}\mathbf{C}^{-1}\mathbf{G}) \tag{5.4.6}$$

where $\mathbf{C}$ is the diagonal matrix given by

$$C_{ii} = \int \varphi_i(s)ds \tag{5.4.7}$$

and $\mathbf{G}$ is the matrix given by

$$G_{ij} = \int \nabla\varphi_i(s)\nabla\varphi_j(s)ds \tag{5.4.8}$$

with $i, j \in \{1, \ldots, 8\}$. According to [36], the vector of weights $\mathbf{x}$ is a Gaussian Markov random field. We provide a detailed explanation in Chapter 4, as well as the proof of equation (5.4.6). We can find in Section 4.5 an explanation on how to calculate the matrices C and G for $D \subset \mathbb{R}^2$. The case of $D \subset \mathbb{R}$ is similar and simpler.

We consider the following model,

$$\begin{aligned} y_i|\mu_i, \sigma_e &\sim N(y_i|\mu_i, \sigma_e) \\ \mu_i &= x_i \\ \mathbf{x}|\tau, k &\sim GF(\mathbf{0}, \mathbf{Q}^{-1}) \,. \end{aligned} \tag{5.4.9}$$

where the predictor $\boldsymbol{\eta}$ is the GF,

$$\eta_i = x_i$$

and the link function $g(\cdot)$ is the identity. However, in our example, the mesh nodes and the locations of the observations are different, so we need to define a predictor $\boldsymbol{\eta}$ for all observations. In order to do so, we define a matrix $\mathbf{A}$, called the *projector matrix*, such that each entry ij is the value of the basis function φ_j at the location s_i,

$$\mathbf{A}_{ij} = \varphi_j(s_i) \,.$$

For example, according to (5.4.5), the row of $\mathbf{A}$ with index $i = 58$ will have the values $\varphi_j(s_{58})$, where $s_{58} = 0.85$. So

$$\begin{aligned} \text{row}^{\mathbf{A}}_{58} &= [\varphi_1(0.85) \;\; \varphi_2(0.85) \;\; \varphi_3(0.85) \;\; \varphi_4(0.85) \;\; \varphi_5(0.85) \;\; \varphi_6(0.85) \;\; \varphi_7(0.85) \;\; \varphi_8(0.85)] \\ &= [0 \;\; 0 \;\; 0 \;\; 0.15 \;\; 0.85 \;\; 0 \;\; 0 \;\; 0] \end{aligned}$$

In this case, the predictor is defined as

$$\boldsymbol{\eta} = \mathbf{A}\mathbf{x} \,,$$

therefore $\boldsymbol{\mu} = \mathbf{A}\mathbf{x}$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$. Let $\tau_e = \frac{1}{\sigma_e^2}$,

$$\pi(\mathbf{y}|\boldsymbol{\eta}, \tau_e, \tau, k) = \prod_{i=1}^{n} \pi(y_i|\eta_i, \tau_e, \tau, k) \propto \prod_{i=1}^{n} \exp\left\{-\frac{\tau_e}{2}(y_i - \eta_i)^2\right\} \,. \tag{5.4.10}$$

On the other hand,

$$\mathbf{x}|\tau, k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1}) \tag{5.4.11}$$

so

$$\pi(\mathbf{x}|\tau, k) \propto |\mathbf{Q}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\mathbf{x}^T \mathbf{Q}\mathbf{x}\right\} \,.$$

The default internal representation of the parameters by `R-INLA` is $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)$ where

$$\theta_0 = \log(\tau_e), \quad \theta_1 = \log(\tau) \quad \text{and} \quad \theta_2 = \log(k) \tag{5.4.12}$$

with independent prior distributions given by

$$\begin{aligned} \theta_0 &\sim \text{LogGamma}(1, 10^{-5}) \,, \\ \theta_1 &\sim \mathcal{N}(0, 1) \,, \\ \text{and} \quad \theta_2 &\sim \mathcal{N}(0, 1) \,. \end{aligned} \tag{5.4.13}$$

In this example, the observations are generated artificially to test the model, as $y(s) = \cos(s)$. To approximate the cosine function, we need to obtain the posterior conditional distribution $\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})$. We have the following,

$$\begin{aligned}\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta}) &\propto \pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\\ &\propto \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \frac{e^{\theta_0}}{2}(\mathbf{y}-\mathbf{A}\mathbf{x})^T(\mathbf{y}-\mathbf{A}\mathbf{x})\right\}\\ &= \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \frac{e^{\theta_0}}{2}\mathbf{y}^T\mathbf{y} + e^{\theta_0}\mathbf{y}^T\mathbf{A}\mathbf{x} - \frac{e^{\theta_0}}{2}(\mathbf{A}\mathbf{x})^T\mathbf{A}\mathbf{x}\right\}\\ &\propto \exp\left\{-\frac{1}{2}\mathbf{x}^T(\mathbf{Q}+e^{\theta_0}\mathbf{A}^T\mathbf{A})\mathbf{x} + e^{\theta_0}\mathbf{y}^T\mathbf{A}\mathbf{x}\right\}. \qquad (5.4.14)\end{aligned}$$

To obtain the mode $\mathbf{x}_0$ of $\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})$, we maximize $\log(\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta}))$. Let

$$\tilde{\mathbf{Q}} := \mathbf{Q} + e^{\theta_0}\mathbf{A}^T\mathbf{A}\,. \qquad (5.4.15)$$

The mode $\mathbf{x}_0$ is the solution of

$$\begin{aligned}&\frac{\partial}{\partial\mathbf{x}}\left(-\frac{1}{2}\mathbf{x}^T\tilde{\mathbf{Q}}\mathbf{x} + e^{\theta_0}\mathbf{y}^T\mathbf{A}\mathbf{x}\right) = 0\\ \Leftrightarrow\quad & -\mathbf{x}^T\tilde{\mathbf{Q}} + e^{\theta_0}\mathbf{y}^T\mathbf{A} = 0^T\\ \Leftrightarrow\quad & \mathbf{x}^T\tilde{\mathbf{Q}} = e^{\theta_0}\mathbf{y}^T\mathbf{A}\\ \Leftrightarrow\quad & \tilde{\mathbf{Q}}\mathbf{x} = e^{\theta_0}\mathbf{A}^T\mathbf{y} \qquad (5.4.16)\\ \Leftrightarrow\quad & \mathbf{x}_0 = e^{\theta_0}\tilde{\mathbf{Q}}^{-1}\mathbf{A}^T\mathbf{y} \qquad (5.4.17)\end{aligned}$$

Consider (5.4.16) and observe the following,

$$\begin{aligned}-\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T\tilde{\mathbf{Q}}(\mathbf{x}-\mathbf{x}_0) &= -\frac{1}{2}\mathbf{x}^T\tilde{\mathbf{Q}}\mathbf{x} + \mathbf{x}_0^T\tilde{\mathbf{Q}}\mathbf{x} - \frac{1}{2}\mathbf{x}_0^T\tilde{\mathbf{Q}}\mathbf{x}_0\\ &= -\frac{1}{2}\mathbf{x}^T\tilde{\mathbf{Q}}\mathbf{x} + e^{\theta_0}\mathbf{A}^T\mathbf{y} - \frac{1}{2}\mathbf{x}_0^T\tilde{\mathbf{Q}}\mathbf{x}_0\end{aligned}$$

and notice that the last term $\mathbf{x}_0^T\tilde{\mathbf{Q}}\mathbf{x}_0$ depends on $\mathbf{y}$ and $\boldsymbol{\theta}$, it does not depend on $\mathbf{x}$. So, from (5.4.14),

$$\begin{aligned}\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta}) &\propto \exp\left\{-\frac{1}{2}\mathbf{x}^T\tilde{\mathbf{Q}}\mathbf{x} + e^{\theta_0}\mathbf{y}^T\mathbf{A}\mathbf{x}\right\}\\ &\propto \exp\left\{-\frac{1}{2}\mathbf{x}^T\tilde{\mathbf{Q}}\mathbf{x} + e^{\theta_0}\mathbf{y}^T\mathbf{A}\mathbf{x} - \frac{1}{2}\mathbf{x}_0^T\tilde{\mathbf{Q}}\mathbf{x}_0\right\}\\ &= \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T\tilde{\mathbf{Q}}(\mathbf{x}-\mathbf{x}_0)\right\}\end{aligned}$$

which is the kernel of a Gaussian distribution with precision matrix

$$\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}} = \mathbf{Q} + e^{\theta_0}\mathbf{A}^T\mathbf{A} \qquad (5.4.18)$$

and mode given by

$$\mathbf{x}_0 = e^{\theta_0}\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}^{-1}\mathbf{A}^T\mathbf{y}\,,$$

therefore,

$$\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{x}_0, \mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}) .$$

To obtain an estimate for $\boldsymbol{\theta}$, first notice that from the property of conditional distributions, we have

$$\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta}) = \pi(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}) = \pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})\pi(\mathbf{y}|\boldsymbol{\theta})$$

so

$$\pi(\mathbf{y}|\boldsymbol{\theta}) = \frac{\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})}{\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})} \tag{5.4.19}$$

therefore, considering (5.4.19) and conditional distributions property,

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) = \frac{\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})}{\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})}\pi(\boldsymbol{\theta}) . \tag{5.4.20}$$

The left side of (5.4.20) does not depend on $\mathbf{x}$, so the right side also does not depend on $\mathbf{x}$. This means that we can choose any value for $\mathbf{x}$ to get a proportional expression for $\pi(\boldsymbol{\theta}|\mathbf{y})$.

In the case of non-Gaussian observations, the distribution $\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})$ is substituted for a Gaussian approximation $\tilde{\pi}_G(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})$, obtained by the Laplace method, explained previously. Then, the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{y})$ is approximated in the following way,

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\mathbf{y}|\mathbf{x}_0,\boldsymbol{\theta})\pi(\mathbf{x}_0|\boldsymbol{\theta})}{\tilde{\pi}_G(\mathbf{x}_0|\mathbf{y},\boldsymbol{\theta})}\pi(\boldsymbol{\theta})$$

where $\mathbf{x}_0$ is the mode of $\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})$.

In our example, the observations are considered to be Gaussian, so $\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})$ is also Gaussian, and instead of an approximation, we have the proportionality given by (5.4.20), where we can choose any value for x. Considering (5.4.13) and (5.4.20), we obtain

$$\begin{aligned}\pi(\boldsymbol{\theta}|\mathbf{y}) \propto (e^{\theta_0})^{\frac{n}{2}} \exp\left\{-\frac{e^{\theta_0}}{2}(\mathbf{y}-\mathbf{A}\mathbf{x})^T(\mathbf{y}-\mathbf{A}\mathbf{x})\right\}|\mathbf{Q}|^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}\right\}\cdot \\ \cdot|\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}|^{-\frac{1}{2}}\exp\left\{\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}(\mathbf{x}-\mathbf{x}_0)\right\}\cdot \\ \cdot 10^{-5}e^{\theta_0}\exp\left\{-10^{-5}e^{\theta_0}\right\}\exp\left\{-\frac{\theta_1^2}{2}\right\}\exp\left\{-\frac{\theta_2^2}{2}\right\} .\end{aligned}$$

Considering that the mode $\mathbf{x}_0$ verifies (5.4.16), and choosing $\mathbf{x} = \mathbf{0}$,

$$\begin{aligned}\pi(\boldsymbol{\theta}|\mathbf{y}) \propto e^{\frac{n\theta_0}{2}}\exp\left\{-\frac{e^{\theta_0}}{2}\mathbf{y}^T\mathbf{y}\right\}|\mathbf{Q}|^{\frac{1}{2}}\;|\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}|^{-\frac{1}{2}}\exp\left\{\frac{e^{\theta_0}}{2}\mathbf{x}_0^T\mathbf{A}^T\mathbf{y}\right\}\cdot \\ \cdot e^{\theta_0}\exp\left\{-10^{-5}e^{\theta_0}\right\}\exp\left\{-\frac{\theta_1^2}{2}\right\}\exp\left\{-\frac{\theta_2^2}{2}\right\}\end{aligned} \tag{5.4.21}$$

Applying the logarithm,

$$\begin{aligned}\log(\pi(\boldsymbol{\theta}|\mathbf{y})) &\propto \frac{n\theta_0}{2} + \frac{1}{2}\log|\mathbf{Q}| - \frac{1}{2}\log|\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}| + \frac{e^{\theta_0}}{2}(\mathbf{A}\mathbf{x}_0-\mathbf{y})^T\mathbf{y} + \theta_0 - 10^{-5}e^{\theta_0} - \frac{\theta_1^2}{2} - \frac{\theta_2^2}{2} \\ &= \frac{1}{2}\log|\mathbf{Q}| - \frac{1}{2}\log|\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}| + \frac{e^{\theta_0}}{2}(\mathbf{A}\mathbf{x}_0-\mathbf{y})^T\mathbf{y} + \left(\frac{n}{2}+1\right)\theta_0 - 10^{-5}e^{\theta_0} - \frac{\theta_1^2}{2} - \frac{\theta_2^2}{2} \\ &\propto \log|\mathbf{Q}| - \log|\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}| + e^{\theta_0}(\mathbf{A}\mathbf{x}_0-\mathbf{y})^T\mathbf{y} + (n+2)\theta_0 - 2\times 10^{-5}e^{\theta_0} - \theta_1^2 - \theta_2^2 .\end{aligned} \tag{5.4.22}$$

The approximate maximum likelihood estimate for $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta}^* \approx \text{argmax}_{\boldsymbol{\theta}}\ \pi(\boldsymbol{\theta}|\mathbf{y})\,. \tag{5.4.23}$$

The expression in equation (5.4.22) was implemented in R, and optimized using the instruction `optim`, independently of the INLA package. The purpose was to get a deep knowledge of the procedure, so we can apply it to different situations, namely, to a different SPDE, with a possibly different precision matrix.

The summary statistics of the posterior distributions of the parameters $(\tau_e, \theta_1, \theta_2)$ can be consulted in Table 5.1.

Table 5.1: Summary statistics for $(\tau_e, \theta_1, \theta_2)$, one dimensional example.

	mean	**st dev**	2.5%**quantile**	50%**quantile**	97.5% **quantile**	**mode**
τ_e	1472.15	213.71	1087.21	1461.01	1924.89	1442.58
θ_1	-0.22	0.34	-0.90	-0.22	0.43	-0.19
θ_2	-0.07	0.38	-0.80	-0.07	0.69	-0.09

We consider the estimates

$$\theta_0^* = \log(1472.15) = 7.29, \quad \theta_1^* = -0.22 \quad \text{and} \quad \theta_2^* = -0.07$$

in (5.4.17) and (5.4.18), and we conclude that the posterior distribution $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ of the GMRF $\mathbf{x}$, is a Gaussian distribution with mode given by

$$\mathbf{x}_0 = (-0.42, 0.56, 0.88, 1.06, 0.59, 0.07, -0.46, -1.06)$$

and precision matrix $\mathbf{Q}_{\mathbf{x}|\mathbf{y},\theta}$ given by

$$\begin{bmatrix} 10573.66 & 4891.97 & 2.22 & 0 & 0 & 0 & 0 & 0 \\ 4891.97 & 14789.92 & 2408.39 & 6.66 & 0 & 0 & 0 & 0 \\ 2.22 & 2408.39 & 9910.08 & 2408.39 & 2.22 & 0 & 0 & 0 \\ 0 & 6.66 & 2408.39 & 14789.37 & 4890.30 & 2.22 & 0 & 0 \\ 0 & 0 & 2.22 & 4890.30 & 14789.37 & 2408.39 & 6.66 & 0 \\ 0 & 0 & 0 & 2.22 & 2408.39 & 9910.08 & 2408.39 & 2.22 \\ 0 & 0 & 0 & 0 & 6.66 & 2408.39 & 14789.92 & 4891.97 \\ 0 & 0 & 0 & 0 & 0 & 2.22 & 4891.97 & 10573.66 \end{bmatrix}.$$

The fitted values for the observations are given by $\hat{\mathbf{y}} = \mathbf{A}\mathbf{x}_0$, where $\mathbf{A}$ is the projector matrix. Figure 5.2 presents the plot of fitted values $\hat{\mathbf{y}}$ (in dashed blue), against observed values $\mathbf{y}$ (in black). The red circled dots are the mode values $\mathbf{x}_0$, located on the mesh nodes, $\{-2, -1, -0.5, 0, 1, 1.5, 2, 3\}$.

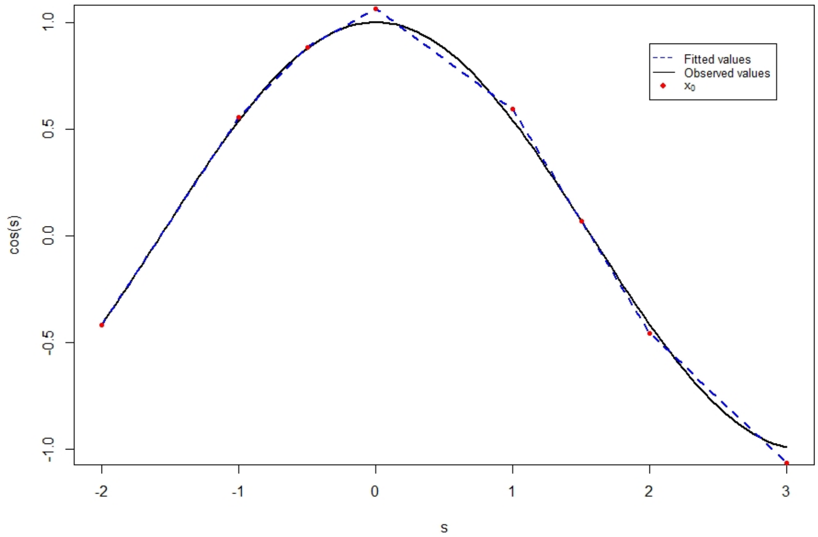


Figure 5.2: Approximation of $\cos(s)$, using the FEM.

5.5 Two dimensional example

Let $\mathbf{y} = (y_1, \ldots, y_n)$ be a vector of n observations at locations $s_i \in D \subset \mathbb{R}^2$. Consider the following model, now with intercept β,

$$\begin{aligned} \mathbf{y}|\beta, \mathbf{x}, \sigma_e^2 &\sim N(\boldsymbol{\mu}, \sigma_e^2 \boldsymbol{I}) \\ \boldsymbol{\mu} &= \mathbf{1}\beta + \mathbf{A}\mathbf{x} \\ \mathbf{x} &\sim GF(0, \mathbf{Q}^{-1}) \end{aligned} \tag{5.5.1}$$

where $\mathbf{A}$ is the projector matrix, $\mathbf{x}$ is a GMRF, and $\mathbf{1}$ is a vector of ones. In this case, the predictor is

$$\eta = \mathbf{1}\beta + \mathbf{A}\mathbf{x} \, .$$

Let $\tau_e = \frac{1}{\sigma_e^2}$ be the precision of the observations, and m be the number of nodes in the mesh, meaning that $\dim(\mathbf{x}) = m$.

It is common to have several zero columns in the projector matrix $\mathbf{A}$. Those columns are related to triangles with no observations: $\text{column}_j^{\mathbf{A}}$ is a zero column if and only if there is no observation y_i in $\text{supp}(\varphi_j)$. Those columns can be dropped, as well as the corresponding $\mathbf{x}_j$ entry of the GMRF $\mathbf{x}$. This may improve the computational calculations.

First we implement formulas (4.5.5), (4.5.6), and (4.5.7), to calculate the matrices $\mathbf{C}$ and $\mathbf{G}$, needed to obtain the Matérn covariance matrix $\mathbf{Q}$ of the GMRF $\mathbf{x}$, which is given by formula (5.4.6).

We have the following,

$$\begin{aligned}\pi(\mathbf{y}|\eta,\tau_e,\tau,k) &= \prod_{i=1}^{n} \pi(y_i|\eta_i,\tau_e,\tau,k) \\ &\propto \prod_{i=1}^{n} \exp\left\{-\frac{\tau_e}{2}(y_i-\eta_i)^2\right\} \\ &= \prod_{i=1}^{n} \exp\left\{-\frac{\tau_e}{2}\left(y_i-\beta-\sum_{j=1}^{m} A_{ij}x_j\right)^2\right\}.\end{aligned}$$

where $x_j = x(s_j)$. On the other hand,

$$\mathbf{x}|\tau,k \sim \mathcal{N}(0,\mathbf{Q})$$

where $\mathbf{Q} = \tau^2(k^4\mathbf{C} + 2k^2\mathbf{G} + \mathbf{G}\mathbf{C}^{-1}\mathbf{G})$, so

$$\pi(\mathbf{x}|\tau,k) \propto |\mathbf{Q}|^{1/2} \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}\right\}.$$

Again, recall the default internal representation of the parameters by `R-INLA`,

$$\boldsymbol{\theta} = (\theta_0,\theta_1,\theta_2) = (\log(\tau_e),\log(\tau),\log(k))$$

with independent *a priori* distributions given by (5.4.13). Moreover, we attribute to the intercept β an *a priori* uniform distribution, $\pi(\beta) \propto 1$. For simplicity of notation, let

$$\boldsymbol{\theta} = (\theta_0,\theta_1,\theta_2,\beta).$$

We have the following,

$$\begin{aligned}\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta}) &\propto \pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\ &\propto \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \frac{e^{\theta_0}}{2}(\mathbf{y}-(\mathbf{1}\beta+\mathbf{A}\mathbf{x}))^T(\mathbf{y}-(\mathbf{1}\beta+\mathbf{A}\mathbf{x}))\right\} \\ &= \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \frac{e^{\theta_0}}{2}\mathbf{y}^T\mathbf{y} + e^{\theta_0}\mathbf{y}^T(\mathbf{1}\beta+\mathbf{A}\mathbf{x}) - \frac{e^{\theta_0}}{2}(\mathbf{1}\beta+\mathbf{A}\mathbf{x})^T(\mathbf{1}\beta+\mathbf{A}\mathbf{x})\right\} \\ &\propto \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + e^{\theta_0}\mathbf{y}^T(\mathbf{1}\beta+\mathbf{A}\mathbf{x}) - e^{\theta_0}(\mathbf{1}\beta)^T\mathbf{A}\mathbf{x} - \frac{e^{\theta_0}}{2}(\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x})\right\} \\ &\propto \exp\left\{-\frac{1}{2}\mathbf{x}^T\left(\mathbf{Q}+e^{\theta_0}\mathbf{A}^T\mathbf{A}\right)\mathbf{x} + e^{\theta_0}(\mathbf{y}-\mathbf{1}\beta)^T\mathbf{A}\mathbf{x}\right\}.\end{aligned}$$

Let $\tilde{\mathbf{Q}} = \mathbf{Q}+e^{\theta_0}\mathbf{A}^T\mathbf{A}$. In order to obtain the mode $\mathbf{x}_0$ of $\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})$, we maximize, for instance, $\log(\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta}))$,

$$\begin{aligned}&\frac{\partial}{\partial\mathbf{x}}\left(-\frac{1}{2}\mathbf{x}^T\tilde{\mathbf{Q}}\mathbf{x} + \beta e^{\theta_0}\mathbf{1}^T\mathbf{y} + e^{\theta_0}(\mathbf{y}-\mathbf{1}\alpha)^T\mathbf{A}\mathbf{x}\right) = 0 \\ &\Leftrightarrow -\mathbf{x}^T\tilde{\mathbf{Q}} + e^{\theta_0}(\mathbf{y}-\mathbf{1}\beta)^T\mathbf{A} = 0^T \\ &\Leftrightarrow \mathbf{x}^T\tilde{\mathbf{Q}} = e^{\theta_0}(\mathbf{y}-\mathbf{1}\alpha)^T\mathbf{A} \qquad (5.5.2) \\ &\Leftrightarrow \mathbf{x}_0 = e^{\theta_0}\tilde{\mathbf{Q}}^{-1}(\mathbf{y}-\mathbf{1}\beta)^T\mathbf{A} \qquad (5.5.3)\end{aligned}$$

Notice the similarity between formulas (5.4.17) and (5.5.3). Considering (5.5.2), observe the following,

$$\begin{aligned} -\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T\tilde{\mathbf{Q}}(\mathbf{x}-\mathbf{x}_0) &= \mathbf{x}^T\tilde{\mathbf{Q}}\mathbf{x} + x_0^T\tilde{\mathbf{Q}}\mathbf{x} + \mathbf{x}_0^T\tilde{\mathbf{Q}}\mathbf{x}_0 \\ &= \mathbf{x}^T\tilde{\mathbf{Q}}\mathbf{x} + e^{\theta_0}(\mathbf{y}-\mathbf{1}\beta)^T\mathbf{A}\mathbf{x} + \mathbf{x}_0^T\tilde{\mathbf{Q}}\mathbf{x}_0 \end{aligned}$$

and notice that $\mathbf{x}_0^T\tilde{\mathbf{Q}}\mathbf{x}_0$ does not depend on x. Therefore,

$$\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T\tilde{\mathbf{Q}}(\mathbf{x}-\mathbf{x}_0)\right\} \tag{5.5.4}$$

which is (again) the kernel of a Gaussian distribution with precision matrix

$$\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}} = \mathbf{Q} + e^{\theta_0}\mathbf{A}^T\mathbf{A} \tag{5.5.5}$$

and mode given by

$$\mathbf{x}_0 = e^{\theta_0}\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}(\mathbf{y}-\mathbf{1}\beta)^T\mathbf{A}\ , \tag{5.5.6}$$

that is

$$\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{x}_0, \mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}})\ .$$

To obtain an estimate for $\boldsymbol{\theta}$, we need to maximize $\pi(\boldsymbol{\theta}|\mathbf{y})$. Applying (5.4.20),

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) = \frac{\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})}{\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})}\pi(\boldsymbol{\theta}) \tag{5.5.7}$$

so we have

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{y}) \propto (e^{\theta_0})^{\frac{n}{2}} \exp\left\{-\frac{e^{\theta_0}}{2}(\mathbf{y}-\mathbf{A}\mathbf{x})^T(\mathbf{y}-\mathbf{A}\mathbf{x})\right\} |\mathbf{Q}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}\right\}\cdot \\ \cdot |\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}|^{-\frac{1}{2}} \exp\left\{\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}(\mathbf{x}-\mathbf{x}_0)\right\}\cdot \\ \cdot 10^{-5}e^{\theta_0}\exp\left\{-10^{-5}e^{\theta_0}\right\}\exp\left\{-\frac{\theta_1^2}{2}\right\}\exp\left\{-\frac{\theta_2^2}{2}\right\} \end{aligned}$$

As said previously, the left hand side of (5.5.7) does not depend on $\mathbf{x}$, so the right hand side also does not depend on $\mathbf{x}$. This means we can choose any suitable value for $\mathbf{x}$. Considering that the mode $\mathbf{x}_0$ satisfies equation (5.5.2), and choosing $\mathbf{x} = 0$, we obtain

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{y}) \propto (e^{\theta_0})^{\frac{n}{2}} \exp\left\{-\frac{e^{\theta_0}}{2}(\mathbf{y}-\mathbf{1}\beta)^T(\mathbf{y}-\mathbf{1}\beta)\right\} |\mathbf{Q}|^{\frac{1}{2}} |\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}|^{\frac{e^{\theta_0}}{2}} \exp\left\{\frac{e^{\theta_0}}{2}(\mathbf{y}-\mathbf{1}\beta)^T\mathbf{A}\mathbf{x}_0\right\}\cdot \\ \cdot e^{\theta_0}\exp\left\{-10^{-5}e^{\theta_0}\right\}\exp\left\{-\frac{\theta_1^2}{2}\right\}\exp\left\{-\frac{\theta_2^2}{2}\right\} \end{aligned}$$

Applying the logarithm,

$$\begin{aligned} \log(\pi(\boldsymbol{\theta}|\mathbf{y})) &\propto \frac{n\theta_0}{2} + \frac{e^{\theta_0}}{2}(\mathbf{A}\mathbf{x}_0 + \mathbf{1}\beta - \mathbf{y})^T(\mathbf{y}-\mathbf{1}\alpha) + \frac{1}{2}\log|\mathbf{Q}| - \frac{1}{2}\log|\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}| + \\ &\quad + \theta_0 - 10^{-5}e^{\theta_0} - \frac{\theta_1^2}{2} - \frac{\theta_2^2}{2} \\ &\propto \log|\mathbf{Q}| - \log|\mathbf{Q}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}| + e^{\theta_0}(\mathbf{A}\mathbf{x}_0 + \mathbf{1}\beta - \mathbf{y})^T(\mathbf{y}-\mathbf{1}\beta) + \\ &\quad + (n+2)\theta_0 - 2\times 10^{-5}e^{\theta_0} - \theta_1^2 - \theta_2^2\ . \end{aligned} \tag{5.5.8}$$

The approximate maximum likelihood estimate for $\boldsymbol{\theta}$ is

$$\boldsymbol{\theta}^* \approx \mathrm{argmax}_{\boldsymbol{\theta}}\ \pi(\boldsymbol{\theta}|\mathbf{y})\ . \tag{5.5.9}$$

In the next section, we consider model (5.5.1), and apply the results of this section to a wind data set describing the velocity of the wind.

5.6 Application of the INLA methodology to the prediction of the wind velocity.

This section contains generated, and also modified, Copernicus Climate Change Service information data set (2021). The H. Hersbach et al. (2018) was downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store. The results contain modified Copernicus Climate Change Service information 2021. It should be mentioned that the European Commission and the European Centre for Medium-Range Weather Forecasts is responsible for any use that may be made of the Copernicus information or data it contains.

We consider the data set H. Hersbach et al. (2018) (see [26]), namely the wind direction components v_x and v_y, at the 1$^{\text{st}}$ of January, 2021, at UTC+0. We consider 500 observations chosen randomly from a total of 4015 from that data set, located at $\{s_i : i = 1, \ldots, 500\}$, with longitude and latitude in the interval $[-15°, 5°]$ and $[30°, 45°]$, respectively.

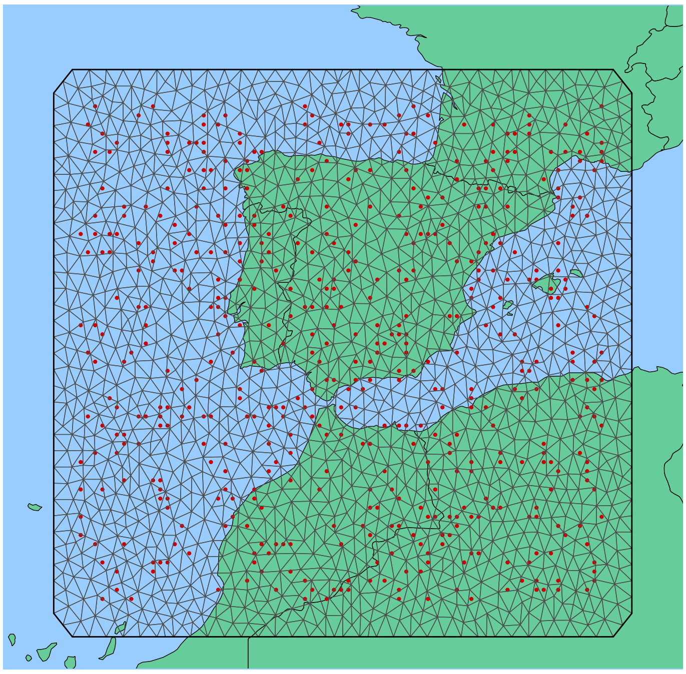


Figure 5.3: Location of the observations (red dots), and mesh.

We use a mesh with 1572 nodes, created by `inla.mesh.2d`, and calculate the matrices **C** and **G** defined by (4.5.5), (4.5.6) and (4.5.7). This matrices allow us to calculate the precision matrix **Q** of the GMRF, **x**. The mesh nodes are different from the locations of the observations, so we need the projector matrix $\mathbf{A}$, given by `inla.spde.make.A`, to calculate (5.5.5) and (5.5.6), precision and mode of the posterior distribution of the GMRF, $\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})$. We can see in Figure 5.3 the locations of the observations (red dots) and the mesh. We consider model (5.5.1), and apply `INLA` for both wind components v_x and v_y, and wind intesity $v = \sqrt{v_x^2 + v_y^2}$. We can find in Tables 5.2, 5.3, and 5.4, the summary statistics of the posterior distributions of the hyperparameters of the model. In this case, we have estimates for the intercept β, the precision of observations τ_e, the practical range ρ, and the standard deviation σ. For $d = 2$ and $\alpha = 2$, we have

$$\begin{cases} \rho = \dfrac{\sqrt{8}}{k} \\ \sigma^2 = \dfrac{1}{4\pi k^2 \tau^2} \end{cases} \quad \text{so} \quad \begin{cases} k = \dfrac{\sqrt{8}}{\rho} \\ \tau = \dfrac{\rho}{4\sqrt{2\pi}\sigma} \end{cases} . \tag{5.6.1}$$

Recall that

$$\boldsymbol{\theta}^* = (\theta_0^*, \theta_1^*, \theta_2^*, \beta^*) \tag{5.6.2}$$

where

$$\theta_0 = \log(\tau_e), \quad \theta_1 = \log(\tau), \quad \theta_2 = \log(k)\,.$$

Table 5.2: Summary statistics for the wind component v_x.

	mean	**st dev**	2.5%**quantile**	50%**quantile**	97.5% **quantile**	**mode**
τ_e	2.28	0.33	1.72	2.25	3.00	2.18
ρ	4.45	0.56	3.43	4.43	5.64	4.38
σ	3.33	0.33	2.73	3.31	4.01	3.29
β	2.18	0.97	0.24	2.18	4.12	2.19

Table 5.3: Summary statistics for the wind component v_y.

	mean	**st dev**	2.5%**quantile**	50%**quantile**	97.5% **quantile**	**mode**
τ_e	2.93	0.46	2.13	2.90	3.94 2.83	
ρ	6.69	0.96	5.02	6.62	8.79	6.46
σ	3.67	0.42	2.92	3.64	4.57	3.58
β	-2.85	1.61	-6.08	-2.85	0.39	-2.85

Table 5.4: Summary statistics for the wind intensity v.

	mean	**st dev**	2.5%**quantile**	50%**quantile**	97.5% **quantile**	**mode**
τ_e	2.43	0.33	1.83	2.41	3.14	2.37
ρ	4.41	0.55	3.47	4.35	5.63	4.23
σ	3.30	0.31	2.75	3.28	3.98	3.22
β	5.99	0.94	4.12	5.99	7.86	5.99

Considering the mean values of (5.6.2), then we calculate fitted values $\hat{\mathbf{y}}$ which are given by

$$\hat{\mathbf{y}} = \mathbf{A}\mathbf{x}_0 + \mathbf{1}\hat{\beta}$$

where $\mathbf{x}_0$ is the mode of the distribution $\pi(\mathbf{x}|\mathbf{y},\boldsymbol{\theta})$ (5.4.17), $\mathbf{1}$ is a vector of ones, and $\hat{\beta}$ is the mean value estimate for β.

The next step is to consider the problem of predicting the expected value of the outcome, on a finer grid of the domain, given by the 4015 localizations from the data set H. Hersbach et al. (2018). We first calculate the projector matrix $\mathbf{A}_g$ from the initial mesh to the new grid executing the instruction `inla.mesh.projector`. Then, the mean of the Gaussian field and predicted values are given by

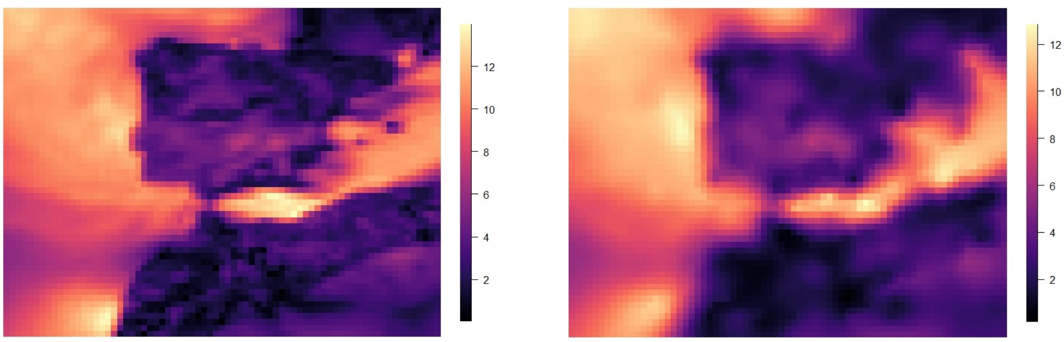
$$\hat{\mathbf{x}}_g = \mathbf{A}_g\mathbf{x}_0 \qquad \text{and} \qquad \hat{\mathbf{y}}_g = \mathbf{A}_g\mathbf{x}_0 + \mathbf{1}\hat{\beta}\,,$$

respectively. Moreover,

$$\mathrm{Var}(\hat{\mathbf{y}}_g) = \mathbf{A}_g\,\mathrm{Var}(\mathbf{x})\,\mathbf{A}_g^T = \mathbf{A}_g\,\mathbf{Q}^{-1}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}\,\mathbf{A}_g^T$$

so the standard deviation of both Gaussian field and predicted values, are calculated as

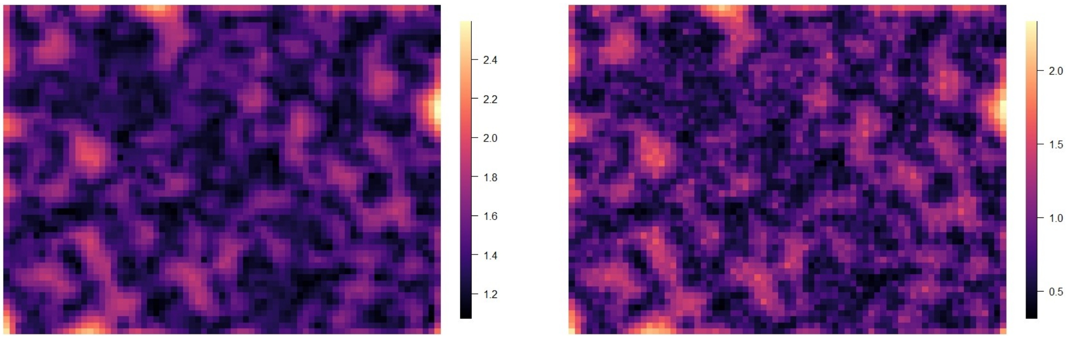
$$\sigma_{\hat{\mathbf{x}}_g} = \mathbf{A}_g\,\mathtt{sqrt}(\mathtt{diag}(\mathbf{Q}^{-1}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}})) \quad \text{and} \quad \sigma_{\hat{\mathbf{y}}_g} = \mathtt{sqrt}(\mathtt{diag}(\mathbf{A}_g\,\mathbf{Q}^{-1}_{\mathbf{x}|\mathbf{y},\boldsymbol{\theta}}\,\mathbf{A}_g^T)).$$



(a) Wind intensity from data set. (b) Predicted wind intensity.

Figure 5.4: Wind intensity v: H. Hersbach et al. (2018) data set versus prediction on a finer grid.

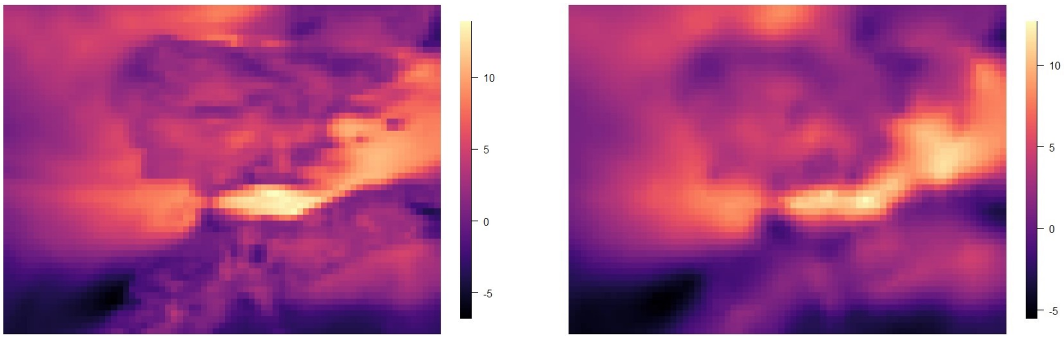
We can find in Figure 5.4 the predicted mean of the outcome $\hat{\mathbf{y}}_g$ (left) versus the "observed values" from the data set (right), on the finer grid, for the wind intensity. The model approximates well the values of the data set, although we can point out that predicted values present a smoother behaviour, resulting of the neighbouring dependence introduced by the Matérn covariance. The model could be improved allowing for another sources of disturbance. We could add fixed effects, for instance, considering pressure and/or temperature as a covariate. The standard deviation values for the predicted Gaussian field and predicted values are represented in Figure 5.5(a) and Figure 5.5(b), respectively.



(a) Standard deviation of the predicted Gaussian field (b) Standard deviation of the predicted wind intensity.

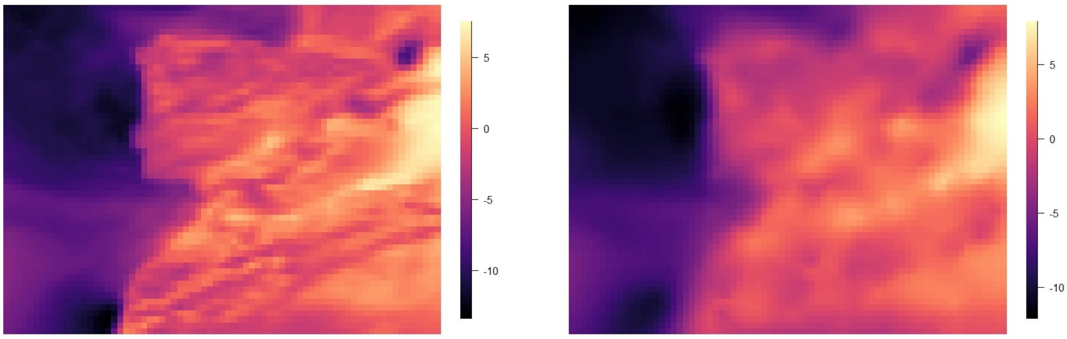
Figure 5.5: Standard deviation of the predicted Gaussian field and predicted wind intensity.

We can find in Figures 5.6 and 5.7 the predicted wind velocity components, v_x and v_y, versus the H. Hersbach et al. (2018) data set. Again, we notice the smoother behaviour of the predicted values.



(a) v_x from data set. (b) Predicted v_x.

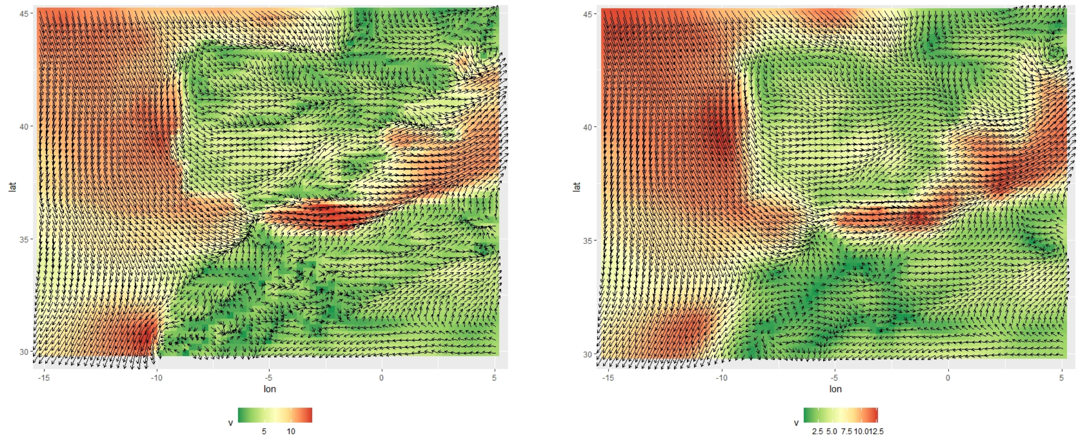
Figure 5.6: v_x component: H. Hersbach et al. (2018) data set versus prediction on a finer grid.



(a) v_y from data set. (b) Predicted v_y.

Figure 5.7: v_y component: H. Hersbach et al. (2018) data set versus prediction on a finer grid.

Regarding the wind velocity, we can find in Figure 5.8 a plot with predicted values for the wind velocity, versus the corresponding quantities from theH. Hersbach et al. (2018) data set. For instance, the region with $(\text{lon}, \text{lat}) \in [-10°, -1°] \times [30°, 37°]$, and $(\text{lon}, \text{lat}) \in [-8°, 4°] \times [39°, 43°]$, present some turbulence that was not captured. This may be due to the geographic features of the land, or other meteorological characteristics, and it is worthy to concentrate our observations on such regions.



(a) Wind velocity from data set.

(b) Predicted values for the wind velocity.

Figure 5.8: Wind velocity: H. Hersbach et al. (2018) data set versus prediction on a finer grid.

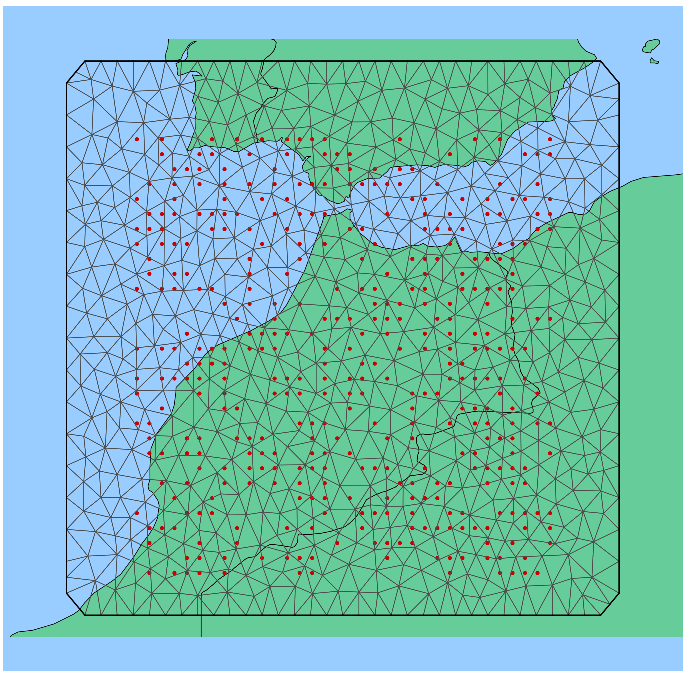


Figure 5.9: Location of the observations (red dots), and mesh.

Therefore we consider 400 observations chosen randomly from a total of 1020 from the data set H. Hersbach et al. (2018), located at $\{s_i : i = 1, \ldots, 400\}$, with longitude and latitude in the interval [−10°, 1°] and [30°, 37°], respectively. We use a mesh with 707 nodes (see Figure 5.9). The results are presented in Tables 5.5, 5.6, and 5.7.

Table 5.5: Summary statistics for the wind component v_x, for the turbulence zone.

	mean	**st dev**	2.5%**quantile**	50%**quantile**	97.5% **quantile**	**mode**
τ_e	5.21	0.70	3.96	5.17	6.70	5.10
ρ	3.83	0.65	2.83	3.73	5.37	3.50
σ	4.18	0.59	3.25	4.09	5.54	3.89
β	1.71	1.73	-1.80	1.71	5.19	1.71

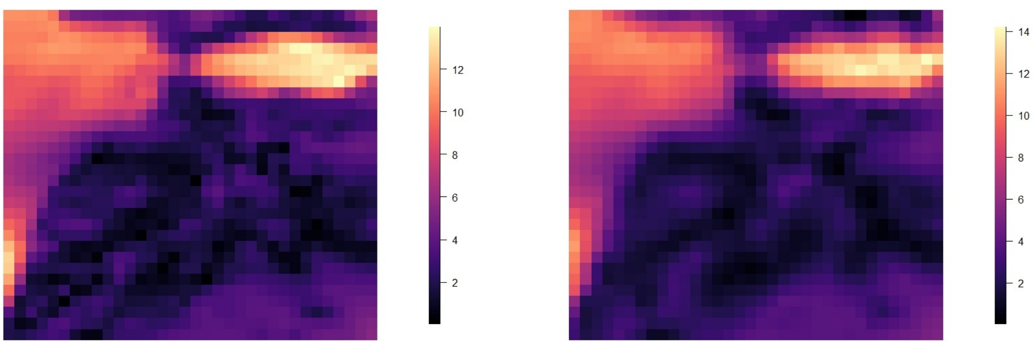
Table 5.6: Summary statistics for the wind component v_y, for the turbulence zone.

	mean	**st dev**	2.5%**quantile**	50%**quantile**	97.5% **quantile**	**mode**
τ_e	3.89	0.59	2.87	3.84	5.19	3.73
ρ	3.51	0.62	2.52	3.42	4.93	3.25
σ	3.08	0.42	2.38	3.03	4.02	2.93
β	-1.83	1.23	-4.33	-1.82	0.63	-1.81

Table 5.7: Summary statistics for the wind intensity v, for the turbulence zone.

	mean	**st dev**	2.5%**quantile**	50%**quantile**	97.5% **quantile**	**mode**
τ_e	4.25	0.56	3.25	4.22	5.45	4.16
ρ	3.86	0.73	2.72	3.75	5.58	3.52
σ	3.96	0.62	2.97	3.88	5.38	3.70
β	4.71	1.73	1.20	4.71	8.25	4.70

We can see in Figure 5.10 the predicted mean for the wind intensity versus the wind intensity from the data set. We conclude that the predicted values capture much better the turbulence of the "observed values".

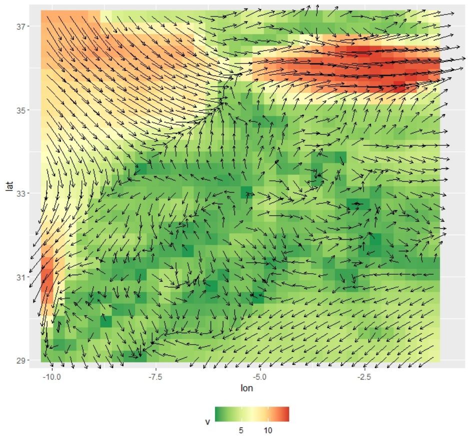


(a) Wind intensity from data set.

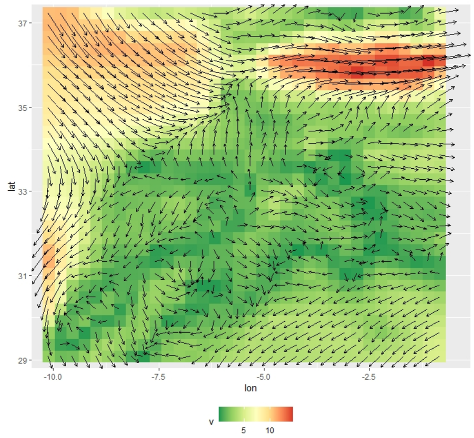
(b) Predicted wind intensity.

Figure 5.10: Wind intensity v: H. Hersbach et al. (2018) data set versus prediction on a finer grid.

In Figure 5.11, we present the predicted wind velocity (left) versus the wind velocity from the data set. Comparing Figures 5.8 and 5.11, we can see that, after concentrating observations in zones with more turbulence, we improve our prediction: the predicted values approximate better the "observed values".



(a) Wind velocity from data set.



(b) Predicted values for the wind velocity.

Figure 5.11: Wind velocity: H. Hersbach et al. (2018) data set versus prediction on a finer grid, for the turbulence zone.

In fact, regarding the estimated practical range ρ for the second data set considered, we can see in Table 5.8 that it decreases for both components v_x, v_y, and also for the wind intensity v, as we should expect, since the zones with more turbulence are regions with a higher heterogeneity of the data, and so we should expect the correlation between neighbouring locations to decrease faster. Therefore, it is useful to concentrate the observations at regions where it is known to exist more turbulence, as the predicted values better approximate the data.

Table 5.8: Pratical range ρ: estimates for the wind velocity components, and intensity.

	v_x	v_y	v
First data set	4.45	6.69	4.41
Second data set	3.83	3.51	3.86

This approach is of great importance, not only because is computationally less expensive than to model complex phenomena numerically, but also because this new statistical approach is supported by a very simple idea, that is, the complexity of the phenomena can be well described by models estimated and updated by data. This is the principle of Bayesian inference, and we can see that it proves to be a quite suitable approach.

6

Final Considerations

In this work we study SPDE, and perform Bayesian inference considering a specific stochastic linear partial differential equation. Then, we apply inference Bayesian, through the INLA methodology, to estimate wind intensity and velocity using a wind data set, H. Hersbach et al. (2018) ([26]).

First we prove the existence and uniqueness of the solution of the stochastic third grade fluid equation,

$$d(v(Y)) = \Big(-\nabla p + \nu\Delta Y - (Y\cdot\nabla)v - \sum_j v^j \nabla Y^j + (\alpha_1+\alpha_2)\mathrm{div}\Big(A^2\Big)$$
$$+ \beta\mathrm{div}\Big(|A|^2 A\Big) + U\Big)dt + \sigma(t,Y)\,d\mathcal{W}_t,$$

with initial conditions in the Sobolev space H^2, and a Navier slip boundary condition. This result was published in [14].

Following [29], the study of spatial data led us to the study of the stochastic linear partial differential equation

$$(k^2-\Delta)^{\alpha/2}\tau x = \mathcal{W}$$

where $\mathcal{W}$ is a Gaussian white noise. We present the mathematical concepts and results that allow us to verify that the solution $x(s)$ is a Matérn field.

Applying the FEM, we are able to construct a finite representation of $x(s)$ of the form

$$\tilde{x}(s) = \sum_{i=1}^{m} x_i \varphi_i(s)$$

where $\mathbf{x} = (x_1,\cdots,x_m)$ is a Gaussian random vector of weights, and $\{\varphi_i : i = 1,\ldots m\}$ is a set of finite element basis functions. Furthermore, the vector $\mathbf{x}$ presents Markov properties, so it is a Gaussian Markov random field. In [29] the authors prove that the finite representation $\tilde{x}(s)$ approximates indeed the Gaussian field $x(s)$.

Then we perform Bayesian inference to specific hierarchical models, and apply the INLA methodology, to obtain an approximated posterior distribution of the parameters of the model.

Finally, we consider that a certain phenomena has a spatial structure, for instance, the wind velocity components, as well as the wind intensity. Under a Bayesian framework, we assume that the spatial structure $x(s)$ is a GF, and collect observations of that phenomena, $\mathbf{y} = \{y_i \in D \in \mathbb{R}^d : i = 1, \ldots, n\}$. The collected data is assumed to be observations of the spatial effect $x(s)$ with an error. Then we use hierarchical Bayesian models for inference, mainly following [37], to update our knowledge of the spatial structure, and get a posterior distribution of the GMRF $\mathbf{x}$. We emphasize that the complexity of the phenomena is captured by the model fed, or informed, by the observations, summarized by the posterior distributions of the model's parameters.

The results obtained for the wind data set show that this statistical new approach approximates well the wind velocity, specially if we concentrate the observation data on localizations where it is known to exist more turbulence of the data. Nevertheless, there is the possibility of improving the model, considering in the mixed effects model, along with a spatial effect, for instance fixed effects, where other covariates are introduced in the model, namely, pressure, temperature, etc. We should point out that the simplicity of the model is not compromised with the introduction of fixed effects.

Furthermore, we plan to continue our work in this field, following the results in [27], where the authors extend the results in [29] to random vector fields, constructing multivariate Gaussian random fields (GRF) using systems of SPDE,

$$\begin{cases} b_{11}(k_{11}^2 - \Delta)^{\alpha_{11}/2} x + b_{12}(k_{12}^2 - \Delta)^{\alpha_{12}/2} x = f_1 \\ b_{21}(k_{21}^2 - \Delta)^{\alpha_{21}/2} x + b_{22}(k_{22}^2 - \Delta)^{\alpha_{22}/2} x = f_2 \end{cases}$$

where f_1 and f_2 are noise processes. GRF are approximated by GMRF with sparse precision matrices, which again allow great computational benefits, when compared to existing multivariate GRF models. This way, instead of applying the INLA methodology to each wind component, we can model both components together. According to [27], some assumptions should be made, in order to keep the model computationally feasible. Nevertheless, it opens the way for a new approach regarding the modelling of complex phenomena.

Bibliography

[1] R. A. Adams and J. J. F. Fournier. *Sobolev spaces*. Elsevier, 2003 (cit. on pp. 6, 8).

[2] S. Albeverio, Z. Brzeźniak, and J.-L. Wu. "Existence of global solutions and invariant measures for stochastic differential equations driven by Poisson type noise with non-Lipschitz coefficients". In: *Journal of Mathematical Analysis and Applications* 371.1 (2010), pp. 309–322 (cit. on p. 23).

[3] A. Bensoussan. "Stochastic navier-stokes equations". In: *Acta Applicandae Mathematica* 38.3 (1995), pp. 267–304 (cit. on p. 2).

[4] A. Bensoussan and R. Temam. "Equations stochastiques du type Navier-Stokes". In: *Journal of Functional Analysis* 13.2 (1973), pp. 195–222 (cit. on p. 2).

[5] M. Blangiardo and M. Cameletti. *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons, 2015 (cit. on pp. 3, 13, 49, 55).

[6] H. Breckner. "Approximation and optimal control of the stochastic Navier-Stokes equation". In: *Mathematisch Naturwissenschaftlich Technischen Fakultät der Martin Luther Universität Halle Wittenberg, Diss* (1999) (cit. on pp. 2, 15).

[7] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer, 2011 (cit. on p. 8).

[8] A. V. Busuioc and D. Iftimie. "A non-Newtonian fluid with Navier boundary conditions". In: *Journal of Dynamics and Differential Equations* 18.2 (2006), pp. 357–379 (cit. on pp. 1, 2, 15, 18–20, 29).

[9] A. V. Busuioc and T. S. Ratiu. "The second grade fluid and averaged Euler equations with Navier-slip boundary conditions". In: *Nonlinearity* 16.3 (2003), pp. 1119–1149 (cit. on p. 22).

[10] V. Busuioc and D. Iftimie. "Global existence and uniqueness of solutions for the equations of third grade fluids". In: *International Journal of Non-Linear Mechanics* 39.1 (2004), pp. 1–12 (cit. on pp. 1, 2, 15, 20, 32, 39).

[11] N. Chemetov and F. Cipriano. "Optimal control for two-dimensional stochastic second grade fluids". In: *Stochastic Processes and their Applications* 128.8 (2018), pp. 2710–2749 (cit. on pp. 18, 23).

[12] N. Chemetov and F. Cipriano. "Well-posedness of stochastic second grade fluids". In: *Journal of Mathematical Analysis and Applications* 454.2 (2017), pp. 585–616 (cit. on pp. 2, 15).

[13] F. Cipriano and A. Cruzeiro. "Navier-Stokes equation and diffusions on the group of homeomorphisms of the torus". In: *Communications in mathematical physics* 275.1 (2007), pp. 255–269 (cit. on p. 2).

[14] F. Cipriano, P. Didier, and S. Guerra. "Well-posedness of stochastic third grade fluid equation". In: *Journal of Differential Equations* 285 (2021), pp. 496–535 (cit. on pp. 15, 83).

[15] F. Cipriano and D. Pereira. "On the existence of optimal and ϵ- optimal feedback controls for stochastic second grade fluids". In: *Journal of Mathematical Analysis and Applications* 475.2 (2019), pp. 1956–1977 (cit. on p. 2).

[16] N. Cressie and C. K. Wikle. *Statistics for spatio-temporal data*. John Wiley & Sons, 2015 (cit. on p. 13).

[17] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge university press, 2014 (cit. on p. 46).

[18] L. Desvillettes and C. Villani. "On a variant of Korn's inequality arising in statistical mechanics". In: *ESAIM: Control, Optimisation and Calculus of Variations* 8 (2002), pp. 603–619 (cit. on p. 17).

[19] J. Dunn and K. Rajagopal. "Fluids of differential type: critical review and thermodynamic analysis". In: *International Journal of Engineering Science* 33.5 (1995), pp. 689–729 (cit. on pp. 2, 15).

[20] L. C. Evans. *Partial differential equations*. Vol. 19. American Mathematical Soc., 2010 (cit. on pp. 7, 8, 46, 53).

[21] R. Fosdick and K. Rajagopal. "Thermodynamics and stability of fluids of third grade". In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 369.1738 (1980), pp. 351–377 (cit. on pp. 2, 15).

[22] A. E. Gelfand et al. *Handbook of spatial statistics*. CRC press, 2010 (cit. on p. 3).

[23] V. Girault and P.-A. Raviart. *Finite element methods for Navier-Stokes equations: theory and algorithms*. Vol. 5. Springer Science & Business Media, 2012 (cit. on p. 8).

[24] T. Hayat, A. Kara, and E. Momoniat. "Exact flow of a third-grade fluid on a porous wall". In: *International Journal of Non-Linear Mechanics* 38.10 (2003), pp. 1533–1537 (cit. on p. 2).

[25] T. Hayat, F. Shahzad, and M. Ayub. “Analytical solution for the steady flow of the third grade fluid in a porous half space”. In: *Applied Mathematical Modelling* 31.11 (2007), pp. 2424–2432 (cit. on p. 2).

[26] H. Hersbach et al. *ERA5 hourly data on single levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS).* [Online; accessed 15-December-2021]. 2018. DOI: `10.24381/cds.adbb2d47` (cit. on pp. 1, 61, 76, 83).

[27] X. Hu et al. “Multivariate Gaussian random fields using systems of stochastic partial differential equations”. In: *arXiv preprint arXiv:1307.1379* (2013) (cit. on p. 84).

[28] E. Krainski et al. *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. Chapman and Hall/CRC, 2018 (cit. on pp. 49, 53, 67).

[29] F. Lindgren, H. Rue, and J. Lindström. “An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73.4 (2011), pp. 423–498 (cit. on pp. 3, 49, 54, 83, 84).

[30] J. M. Lourenço. *The NOVAthesis LaTeX Template User’s Manual*. NOVA University Lisbon. 2021. URL: `https://github.com/joaomlourenco/novathesis/raw/master/template.pdf` (cit. on p. iii).

[31] B. Murteira, C. Paulino, and M. Turkman. *Estatística Bayesiana*. Fundação Calouste Gulbenkian, 2003 (cit. on p. 11).

[32] A. Rasheed et al. “Stabilized approximation of steady flow of third grade fluid in presence of partial slip”. In: *Results in physics* 7 (2017), pp. 3181–3189 (cit. on p. 2).

[33] P. A. Razafimandimby and M. Sango. “Strong solution for a stochastic model of two-dimensional second grade fluids: existence, uniqueness and asymptotic behavior”. In: *Nonlinear Analysis: Theory, Methods & Applications* 75.11 (2012), pp. 4251–4270 (cit. on pp. 2, 15, 23).

[34] Y. Rozanov. *Random fields and stochastic partial differential equations*. Vol. 438. Springer Science & Business Media, 1998 (cit. on pp. 9–11, 46).

[35] W. Rudin. *Real and complex analysis*. McGraw-Hill, 1987 (cit. on p. 46).

[36] H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC, 2005 (cit. on pp. 3, 49–52, 68).

[37] H. Rue, S. Martino, and N. Chopin. “Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations”. In: *Journal of the royal statistical society: Series b (statistical methodology)* 71.2 (2009), pp. 319–392 (cit. on pp. 3, 4, 56, 62, 64, 84).

[38] P. Whittle. “On stationary processes in the plane”. In: *Biometrika* (1954), pp. 434–449 (cit. on pp. 3, 48).

[39] P. Whittle. “Stochastic processes in several dimensions”. In: *Bulletin de L’Institut International de Statistique* 40 (1963), pp. 974–994 (cit. on pp. 3, 48).

2022 Stochastic Partial Differential Equations and Applications in Bayesian Inference Sílvia Guerra