UNIVERSITY OF SUSSEX

School of Mathematics and Physical Sciences

Department of Mathematics

Some approximation results for Bayesian inverse problems

Stylianos Katsarakis

Supervisors:

Masoumeh Dashti

Charalambos Makridakis

May 2021

Submitted for the degree of Doctor of Philosophy

University of Sussex

Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree. This thesis is comprised entirely of my own research, conducted under the supervision of Masoumeh Dashti. Work due to other authors will be made clear and cited where appropriate.

Signature:

....

Stylianos Katsarakis

Brighton, 2nd May 2021

University of Sussex

School of Mathematical and Physical Sciences

Department of Mathematics

Doctor of Philosophy

Stylianos Katsarakis

Some approximation results for Bayesian inverse problems

Abstract

The current thesis consists of two results obtained during my PhD, both related to approximations of high/infinite-dimensional measures emerging from the Bayesian approach to inverse problems. In the first part, we study a technique for the reduction of the dimension in the finite but high-dimensional case when the prior is 1-exponentially distributed. In Chapter 4, this is done in a way that the approximated posterior measure minimises the distance to the posterior by using an appropriate metric. In the second part, we consider the problem of estimating the drift and diffusion coefficient of a stochastic differential equation using noisy measurements on a single path. There, we use a perturbation technique on the solution of the SDE to obtain an approximated posterior; in Chapter 5, we study the convergence properties of this approximation.

The author was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 642768, ModCompShock.

Contents

List of Abbreviations and Symbols

$a \propto b$	: It means that there exists a constant C such that $a = Cb$.
$a \lesssim b$	: It means that there exists a constant C such that $a \leq Cb$.
f_X, $f_{N(0,\gamma^2)}, f_\pi$	: It denotes a density function of a random variable X, or a probability $N(0, \gamma^2)$, π.
Σ_X	: It is the $\sigma-$algebra of the measurable space (X, Σ_X).
$\mathbb{E}_{X\sim\pi}\left(f(X)\right)$, $\mathbb{E}_\pi\left(f(X)\right)$	: Expectation of $f(X)$, where $X \sim \pi$.
$\sigma(Y)$	: It is the smallest $\sigma-$algebra which is generated by a random variable Y.
$\mathbb{E}\left(f(X)t \middle\vert \sigma\left(Y\right)\right)$	: Conditional expectation of X given $\sigma-$algebra $\sigma(Y)$.
$Ent_\pi\left(f\right)$	: Entropy of $f(X)$, where $X \sim \pi$, see Definition 4.4.1.
$U = (a_1\vert a_2\vert \dots \vert a_n)$	: U is the matrix with columns the vectors a_1, $a_2, \dots$, a_n.
$U = (U_1\vert U_2)$	: The above definition, it also extend also with matrices instead vectors.
$\pi_{m,B}$	: In Chapter 4, we define the one exponential probability measure, where m, B are parameters of the probabilty measure $\pi_{m,B}$.
$L^\infty(D)$	: It is a function space which contains bounded function over the domain D.

Chapter 1

Introduction

In the present thesis, we study the approximation of measures defined on high or infinite dimensional spaces. In particular, these measures emerge in the context of the Bayesian approach to inverse problem.

Let us start with the definition of inverse problem with measurements subject to additive noise. Without noise, the measurement y is generated by the forward model represented by $G : X \to Y$. We hence write

$$y = G(u) + \eta$$

The aim of the problem is to recover the unknown u from the knowledge of y. In general, either the distribution or the size of noise term η is known. In our set-up, we assume that the distribution of η is given.

Intuitively speaking, the Bayesian approach to the inverse problem can be described as follows. Given prior information about the location of the unknown u within X and measurements y, we are interested in an update of our prior information according to the measurements y. In the context of Bayesian approach, both the prior information and its updated version are considered to be distributions over elements of space X, and we call them prior distribution and posterior distribution, respectively. Essentially, this approach emerges from Bayes' Theorem, if both prior and posterior distributions admit a probability density function, Bayes' formula provides us the following expression

$$f_{U|Y}(u|y) \propto f_{Y|U}(y|u) f_U(u)$$

In a more general mathematical framework and under appropriate conditions, Bayes' formula provides us with the Radon-Nikodym derivative of the posterior probability with respect to the prior probability, and the derivative is the so-called likelihood.

In statistics, one needs to introduce quantities of interest for a better understanding of distributions, such examples are the mean value of distribution, the variance, the mode, etc. Returning to the Bayesian approach and considering large or infinite dimensional X, such quantities are usually expected values of bounded continuous functions with respect to the posterior and are approximated using Monte Carlo methods. An advantage of using Monte Carlo is that the convergence rate of those methods is the same regardless the dimensions of the problem.

Let us introduce the Importance Sampling method; this is a method derived from the theory of Monte Carlo simulations. In this concept, we consider the proposal distribution π, and we are interested in evaluating the target distribution μ. An essential condition is that μ needs to be absolute continuous with respect to π and its Radon-Nikodym derivative to be known up to a multiplication constant

$$\frac{d\mu}{d\pi}(x) \propto f(x)$$

Given the structure of the Bayesian approach, i.e. the Radon-Nikodym is known, the Importance Sampling is a suitable method for the approximation of the posterior distribution.

In most presentations of Importance Sampling, it is highlighted that a way to accelerate the computational time of the Monte Carlo is to focus on an appropriate region of X. Essentially, we consider a subset that contributes the most to the evaluation of our Monte Carlo simulation.

Let us now consider the following example. We assume probabilities μ and π are defined on the infinite dimensional space $\mathbb{R}^\infty$, and the Radon Nikodym derivative of μ with respect to π is defined as the infinite product of a non-trivial function g, i.e. $g \not\equiv 1$, then we have that μ and π are singular, see for instance Agapiou et al. (2017). Consequently, we have that for non-trivial probabilities μ and π defined on the infinite dimensional space X with μ absolutely continuous with respect to π, the part of space X, which contributes the most to the evaluation, tends to be restricted to low-dimensional subspaces.

That subspace is derived by a minimisation problem. More precisely, we consider candidate measures for the approximation of the target measure μ. Such measures correspond to a subspace of X and the approximated measure is the one that minimises the distance between the candidate measures and μ. The distance used in the minimisation problem is the Kullback-Leibler divergence. The advantage of this technique is that it applies to the Bayesian approach with nonlinear forward operators, non-Gaussian priors, and non-Gaussian noise term, also provides us with a relation between the considered dimension r and the error of the approximation. This methodology was proposed in Zahm et al. (2021) where the authors consider the case of sub-Gaussian prior distributions.

My contribution to the above mentioned methodology is its extension to the case of 1-exponential measures. In Sections 4.4.1 and 4.4.2, one can find my contributions to that methodology. More precisely, my main contribution is Theorem 4.4.3, which allows to extend the methodology for the case of 1-exponential measure. Also, the results of that methodology are summarised in Proposition 4.4.1. Therein one can find error estimates of the method with respect to Kullback Leibler divergence and the way to recover that optimal space.

As we see in Section 2.3.1, the 1-Besov priors are defined through 1-exponential distributions. These priors are important especially for the field of signal and image processing because of their edge-preserving and sparsity-promoting properties, see e.g. Leporini and Pesquet (2001); Kolehmainen et al. (2012); Jia et al. (2016); Tan Bui-Thanh (2015); Rantala et al. (2006).

In the second part of the thesis, we study the problem of estimating the drift and diffusion coefficients of a stochastic differential equation with a small diffusion coefficient from discrete noisy measurements of a single path of the solution. Such equations are usually represented as,

$$dX(t) = (a(X(t)) - b(X(t))\,)\;dt + \frac{1}{\sqrt{N}}\,\sqrt{a(X(t)) + b(X(t))}\,dW_t \tag{1.1}$$

where N is a large number, i.e. $N \gg 1$. We assume that the process is restricted to the bounded interval $[0,1]$ and measurements are collected from a realisation of $(X\,(t_1)\,,\dots X\,(t_n))$ at times $0 \le t_1 < \dots t_n \le T$, for positive time T.

Let the forward map $G : C^2([0,1]) \times C^2([0,1]) \to \mathbb{R}^n$ be defined as

$$G(a,b) := (X\,(t_1)\,,\dots X\,(t_n))(a,b)$$

where $X\,(t_i)\,(a,b)$ is the solution of the above SDE with coefficients $a-b$ and $\frac{1}{\sqrt{N}}\sqrt{a+b}$. Denote the noisy measurements by $y \in [0,1]^n$ and the noise term by $\epsilon\,\eta = (\epsilon\,\eta_1, \dots \epsilon\,\eta_n)$, where $\epsilon > 0$ and η_i are independent and identically distributed. Then, the above inverse problem can be written in the following form,

$$y = G(a,b) + \eta$$

where the functions a and b are unknown.

At this point it is worth mentioning that equation (1.1) is closely connected to the birth death process. Consider for instance $(Y_t)_{t\ge 0}$ birth death process with state space $\{0,\dots,N\}$, where $N, T > 0$, as described for example in Renshaw (2015). Next, we consider the following two set of parameters: u_k is the up jump rate the k state and d_k

down jump rate, with $k \in \{0, \ldots, N\}$, and an initial condition Y_0. Under the condition that $u_N = d_0 = 0$ and $Y_0 \in \{0, \ldots, N\}$, we have that the process Y_t will remain in $\{0, \ldots, N\}$. Suppose now that these jump rates have the so-called density dependent property Ethier and Kurtz (2009), meaning that there exists functions $a, b : [0, 1] \to \mathbb{R}$ such that:

$$\forall k \in \{0, \ldots, N\}, \frac{u_k}{N} = a\left(\frac{k}{N}\right) \text{ and } \frac{d_k}{N} = b\left(\frac{k}{N}\right),$$

and where we assume that U and D are non-negative on $[0, 1]$. Then, it is well-known, see for instance Kurtz (1971), that in the limit of large N, the rescaled BD process $\left(\frac{1}{N}Y_t\right)_{t\geq 0}$ can be approximated by a diffusion process, that satisfies equation (1.1) with reflective boundary conditions and initial condition $x_0 = \frac{1}{N}Y_0 \in [0, 1]$.

In Chapter 5, we consider the posterior distribution μ^y that emerges from the above inverse problem. More precisely, we have that the likelihood is the product of transition probabilities, which are the fundamental solution of a parabolic differential equation. Such solutions are usually expensive to compute. We propose an approximation of the likelihood using a random perturbation technique. This leads to an approximation of solution X, in the form of $X^N \approx X_0 + \frac{1}{\sqrt{N}}X_1$ with X_0 the solution of a deterministic first order differential equation and X_1 a Gaussian process. As a result, the likelihood of the approximated posterior is much easier to compute. We establish some convergence properties of X^N under appropriate regularity conditions on the drift and diffusion, and study the convergence properties of the resulting approximated posterior.

1.1 Organisation of the thesis

The thesis is organised as follows. Chapter 2 summarises some of the already known background materials for the Bayesian approach to Inverse problems and the Importance Sampling method. In particular, Section 2.5 demonstrates a version of the Bayes' Theorem on Banach Space. Later in Section 2.5, there is a general framework for approximating the posterior based on an approximation of the forward map; this is the same framework that we desire to establish for the approximation of the posterior of the second problem, see above. Later on, in the same chapter, there is an introductory section for Monte Carlo simulations. More precisely, Section 2.6.1 is devoted to the Importance Sampling technique and provides some materials about Kullback-Leibler divergence.

Chapter 3 is an introductory Section for stochastic differential equations, where the reader can find results relevant to the existence and uniqueness of the solution. We also study the behaviour of the solution near the spatial boundary. Furthermore, Section 3.3

is devoted to the Fokker-Planck equation.

In Chapter 4, we first recall the methodology proposed by Zahm et al. (2021) for dimension reduction in the case of sufficiently regular prior. Then, in our main result, we extend this dimension reduction technique to the case of 1-exponential priors in Section 4.4.1.

In Chapter 5, we study the above mentioned inverse problem which involves the stochastic differential equation. Section 5.2 explores the perturbation technique and introduces the ODE and the Gaussian process, which describe X_0 and X_1, respectively. In addition, we demonstrate our results related to the convergence of perturbation technique. In Section 5.3, we build the framework associated with the approximation of the posterior.

Chapter 2

Bayesian inverse problems

This chapter is devoted to background material related to Bayesian approach to inverse problems, and gives an introduction to Monte Carlo methods. In particular, Section 2.1 starts with the generic inverse problem and then focuses on inverse problems with measurements subject to additive noise which are the main interest of the current thesis. Afterward, Section 2.2 presents the Bayesian approach to inverse problems, where we present Bayes' Theorem on Banach spaces, which plays a central role to the definition of the current approach. We then review the conditions for the well-definedness of the solution emerging from the Bayesian framework, the so-called posterior distribution. We then study the effect of the approximation of the underlying forward operator on the posterior distribution. Finally, Section 2.6 is an introduction on Monte Carlo methods. More precisely, it focuses on importance sampling, which is a suitable method for approximating the posterior distribution that emerges from the framework of the Bayesian approach.

2.1 A generic formula for inverse problems

Suppose X, Y are Banach spaces. We consider the problem of recovering the unknown $u \in X$ from some data $y \in Y$ given the model

$$y = \tilde{G}(u),$$

where $\tilde{G} : X \to Y$ is a generic stochastic mapping that associates the unknown variable u with the observed data y. We usually call it the observation map.

One example is the case where a noise effect has been incorporated into the observed data, and the stochastic map $\tilde{G}$ is given as follows

$$\tilde{G}(u) = G(u) + \eta$$

where $G : X \to Y$ is a deterministic map that relates u to y, G is usually called forward map. The noise term η is also known in some sense, one may consider either its distribution or its magnitude to be known. In our set-up, we assume that the distribution of η is given.

Intuitively speaking, we have been asked to invert the operator $\tilde{G}$. One strategy for recovering the unknown u is to deploy an estimator which behaves similarly to the inverse of the operator $\tilde{G}$.

There are several difficulties that one needs to overcome in this problem:

1. Non-existence:

Consider the inverse problem of recovering u from the data

$$y = G(u) + \eta$$

where η is an element in $\mathbb{R}^n$ and the forward map $G : \mathbb{R}^n \to Im(G)\,(\subsetneq \mathbb{R}^n)$ is invertible. Consider the situation where one uses the inverse of G as the estimator which retrieves u. Observe now that since we only know the distribution of η, we have that y and η are inseparable and that there may be several choices of η such that $y \notin Im(G)$, thus there are several y for which our estimator cannot solve our problem.

2. Non-uniqueness:

We consider the inverse problem without the noise term here for simplicity, and assume an underdetermined system of equations, that is a linear map $G : \mathbb{R}^n \to \mathbb{R}^m$, where $n > m$, satisfying

$$y = G(u).$$

In this example, we may have several $u \in \mathbb{R}^n$ solving the above equation.

3. Instability:

Let us once again consider the problem

$$y = G(u) + \eta$$

where G is an invertible linear map $G : \mathbb{R}^n \to \mathbb{R}^n$, also assume that the noise term is bounded by some $\epsilon > 0$, $\|\eta\| \leq \epsilon$, where $\|\cdot\|$ is some norm over $\mathbb{R}^n$. In addition, we consider the specific case, where G is a diagonal matrix and the elements on the diagonal decay fast to zero, for instance $\lambda_i = 1/i$. Then the inverse of G is given as follows

$$G^{-1} = \begin{pmatrix} 1/\lambda_1 & & \\ & \ddots & \\ & & 1/\lambda_n \end{pmatrix}$$

Then using the inverse map as an estimator for the unknown u may mislead us as we see now. Let the noise $\eta = (0, \ldots, 0, \epsilon)$ and $n \gg 1$. Denote the data with noise by y and without noise by $y^\dagger = G(u^\dagger)$. Let us now check the contribution of the error term in the estimation of the unknown u,

$$\|u - u^\dagger\| = \|G^{-1}(y - y^\dagger)\| = \|\,(0, \dots, n\,\epsilon)\,\| = n\,\epsilon\|\,(0, \dots, 1)\,\|.$$

Therefore, we observe that small noise in the data could lead us in large changes of the solution u.

The above difficulties are summarised in Hadamard's definition of well-posedness: A well-posed problem is a mathematical model which has the following three properties:

(a) The solution exists.

(b) The solution is unique.

(c) The solution is stable, i.e. small variations in the observed data cause small changes to the solution.

In the following section, we introduce the Bayesian approach for inverse problems, a probabilistic approach for recovering the unknown u in a wellposed manner. Section 2.4 presents a suitable set of assumptions which addresses the well-posedness of the Bayesian approach.

2.2 The Bayesian approach for inverse problems

Let us consider the inverse problem of finding u from y given as

$$y = G(u) + \eta. \tag{2.1}$$

We assume that (X, Σ_X) and (Y, Σ_Y) are measurable spaces, $G : X \to Y$ is a measurable map and the distribution of η is given and $\eta \in Y$.

Intuitively speaking, this approach can be described as follows: Given prior information about the location of the unknown u within X, and measurements y, we are interested in an update of our prior information according to the measurements y. Both the prior information and its updated version are considered to be distributions over space X, and we call them prior distribution and posterior distribution, respectively.

In this approach, we treat u, y and η as random variables and determine the joint distribution of (u, y). Notice that the distribution of η is given above. Given the prior distribution and knowing that u and η are independent, we can then determine the joint distribution of (u, y) using (2.1). Finally, we obtain the conditional distribution of $u|y$, the so-called posterior distribution in this approach. As we show in the following sections, Bayes' rule plays a central role in obtaining the posterior distribution. If u and y are finite-dimensional random variables, Bayes' rule implies the following for the density functions of the concerned random variables

$$f_{U|Y}(u|y) = \frac{1}{f_Y(y)} f_{Y|U}(y|u) f_U(u). \tag{2.2}$$

In the context of Bayesian approach, the density function $f_{Y|U}(y|u)$ is called likelihood.

Using this approach allows us to incorporate a priori information about the unknown u by choosing an appropriate prior distribution.

Next, we provide some notations to facilitate our study of the distribution of u, η and $u|y$, which are used in the following section. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an abstract probability space and (X, Σ_X) and (Y, Σ_Y) measurable spaces. We consider y and u as realisations of the measurable functions $\mathcal{Y} : \Omega \to Y$ and $\mathcal{U} : \Omega \to X$. We use the notation $\eta : \Omega \to Y$ for the random variable representing noise as there is no room for confusion. Then we define the two probability measures $\mathbb{Q}_0(\cdot) = \mathbb{P}(\eta \in \cdot)$ and $\mu_0(\cdot) = \mathbb{P}(\mathcal{U} \in \cdot)$ over (Y, Σ_Y) and (X, Σ_X), respectively. Observe that according to the same notation the posterior probability can be written as $\mu^y(\cdot) = \mathbb{P}(\mathcal{U} \in \cdot\, |\mathcal{Y} = y)$ defined on Σ_X.

Looking at the Bayes' rule (2.2), we also need to determine the probability emerging from the random variable $\mathcal{Y}|\{\mathcal{U} = u\}$. For any $A \in \Sigma_Y$ and $u \in X$, we have

$$\begin{aligned} \mathbb{Q}_u(A) &:= \mathbb{P}(\mathcal{Y} \in A|\mathcal{U} = u) = \mathbb{P}(G(\mathcal{U}) + \eta \in A|\mathcal{U} = u) = \\ &\mathbb{P}(\eta \in A - G(u)\,|\mathcal{U} = u) = \mathbb{P}(\eta \in A - G(u)) = \mathbb{Q}_0(A - G(u)) \end{aligned} \tag{2.3}$$

where the third equality holds due to the independence of u and η.

Now Bayes' rule (2.2) provides us with the following Radon-Nikodym derivative

$$\frac{d\mu^y}{d\mu_0}(u) = g(u; y) \Big/ \int_X g(x; y)\mu_0(dx).$$

Notice that the Radon-Nikodym derivative is proportional to $g(u; y)$ which is the so-called likelihood, see for instance (2.2).

In the following two sections, we present an example of a finite dimensional space and then we introduce Bayes' theorem defined on Banach spaces.

2.2.1 Application of Bayes' Theorem in finite dimensions

Let us consider two euclidean spaces $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$ equipped with the Borel σ-algebra, measurable map $G : X \to Y$ and given probabilities μ_0 and $\mathbb{Q}_0$ defined as above. We further assume that μ_0 and $\mathbb{Q}_0$ admit a probability density function ρ_0 and ρ, respectively, i.e.

$$\mathbb{Q}_0 (A) = \int_A \rho (\eta) \mathcal{L}^m (d\eta) \qquad A \in \mathcal{B}(\mathbb{R}^m)$$
$$\mu_0 (B) = \int_B \rho_0 (u) \mathcal{L}^n (du) \qquad B \in \mathcal{B}(\mathbb{R}^n)$$

where $\mathcal{L}^n$ is the Lebesgue measure defined on $\mathcal{B}(\mathbb{R}^n)$ and analogously for $\mathcal{L}^m$. We can then define the joint distribution $(\mathcal{U}, \mathcal{Y})$ in terms of $\mathbb{Q}_u$ and μ_0, but first let us observe that the density function of $\mathbb{Q}_u$ is equal to $\rho(y - G(u))$, for fixed u, see for example equation (2.3). Therefore, the density function of $(\mathcal{U}, \mathcal{Y})$ is proportional to $\rho(y - G(u)) \rho_0(u)$. If we further assume that

$$\int_{\mathbb{R}^n} \rho(y - G(u)) \rho_0(u) \mathcal{L}^n(du) > 0^{1}, \qquad \text{for a given } y \in \mathbb{R}^m$$

then Bayes' rule implies the following, see for instance Theorem 1.1 in Dashti and Stuart (2015),

$$\frac{d\mu^y}{d\mathcal{L}^n}(u) \propto \rho(y - G(u)) \frac{d\mu_0}{d\mathcal{L}^n}(u). \tag{2.4}$$

Or, in other words

$$\mu^y(A) := \int_A \frac{\rho_0(y - G(u))}{\int_{\mathbb{R}^n} \rho_0(y - G(\tilde{u})) \mu_0(d\tilde{u})} \mu_0(du).$$

Notice that Bayes' theorem, as it is stated in this section, requires the joint distribution $(\mathcal{U}, \mathcal{Y})$ to admit a probability density function, i.e. Radon-Nikodym derivative with respect to the Lebesgue measure. The following section demonstrates Bayes' theorem on separable Banach spaces. In an infinite dimensional space, the main problem is the absence of a Lebesgue measure.

2.2.2 Bayes' Theorem on Banach Spaces

Suppose that X and Y are separable Banach spaces equipped with their respective Borel σ-algebra, $G : X \to Y$ is a measurable mapping, and that we are interested in solving the inverse problem (2.1).

As we mentioned in the description of the Bayesian approach, the prior distribution and the noise distribution are known, so that we can derive the joint distribution $(\mathcal{U}, \mathcal{Y}) \in X \times Y$ from which we are interested in computing the conditional distribution $\mathcal{U}|\mathcal{Y}$.

Let us consider the same notation as in the last two sections:

- $\mathbb{Q}_0$ and $\mathbb{Q}_u$ are defined on Σ_Y, the distributions of noise and $\mathcal{U}|\mathcal{Y}$, respectively.
- μ_0 and μ^y are defined on Σ_X, prior and posterior distributions, respectively.

[1]Notice that the above integral is the marginal distribution of y, i.e. $\mathbb{P}(y) \neq 0$, see for example (2.2)

Notice also that the σ-algebras of X and Y are given as follows, $\Sigma_X = \mathcal{B}(X)$ and $\Sigma_Y = \mathcal{B}(Y)$. In the following theorem, we assume that $\mathbb{Q}_u$ is absolutely continuous with respect to $\mathbb{Q}_0$ for μ_0-almost every u, i.e.

$$\mathbb{Q}_u \ll \mathbb{Q}_0, \quad \forall\ \mu_0\text{-a.e. } u.$$

According to Theorem 3.8 in Folland (2013), we have that for μ_0-almost every u, there exists a $\mathcal{B}(Y)$-measurable function $\Phi(u,\cdot) : X \times Y \to \mathbb{R} \cup \{+\infty\}$ such that $e^{-\Phi(u,\cdot)} \in L^1(\mathbb{Q}_0)$[2] and

$$\frac{d\mathbb{Q}_u}{d\mathbb{Q}_0}(y) := e^{-\Phi(u,y)}. \tag{2.5}$$

Since $\mathbb{Q}_0$ is assumed to be a probability measure, it also holds that $\mathbb{E}_{y\sim\mathbb{Q}_0} e^{-\Phi(u,y)} = 1$, for μ_0-a.e. u.

As noted above, the approach starts with the determination of the joint distribution of $(\mathcal{U}, \mathcal{Y})$, thus we define probability ν which is defined on $\mathcal{B}(X) \otimes \mathcal{B}(Y)$,

$$\nu(du, dy) := \mathbb{P}(\mathcal{U} \in du,\, \mathcal{Y} \in dy) = \mathbb{Q}_u(dy)\mu_0(du) = \mu^y(du)\mathbb{P}(\mathcal{Y} \in dy).$$

To compensate for the absence of a Lebesgue measure in function spaces, we consider the probability measure ν_0 defined on $\mathcal{B}(X) \otimes \mathcal{B}(Y)$ and given by

$$\nu_0(du, dy) = \mathbb{Q}_0(dy)\mu_0(du)$$

as the reference measure. In what it follows, we assume that $\Phi(\cdot,\cdot)$ is ν_0-measurable. Note that the definition of ν and ν_0 implies that $\nu \ll \nu_0$ and its Radon-Nikodym derivative is written as

$$\frac{d\nu}{d\nu_0}(u, y) = e^{-\Phi(u,y)}.$$

The following theorem summarises all the above assumptions and it is applicable to inverse problem with measurements subject to additive noise, in the case where X and Y are separable Banach spaces. The proof can be found in Dashti and Stuart (2015).

Theorem 2.2.1 (Bayes' Theorem)**.** *Suppose $\Phi : X \times Y \to \mathbb{R}$, given in (2.5), is ν_0-measurable, and*

$$Z(y) := \int_U e^{-\Phi(u,y)} \mu_0(du) \ > 0, \qquad \forall \mathbb{Q}_0 - a.s.\ y.$$

Then the conditional distribution of $\mathcal{U}|\mathcal{Y} = y$, denoted by $\mu^y(\cdot)$, exists under ν. Furthermore, if $\mu^y \ll \mu_0$ for ν-a.s. y, then we have

$$\frac{d\mu^y}{d\mu_0}(u) := \frac{1}{Z(y)} e^{-\Phi(u,y)}. \tag{2.6}$$

[2] $L^1(\mathbb{Q}_0)$ is the space of all real-valued $\mathbb{Q}_0$-integrable functions

2.3 Random series construction of priors

This section presents how to build a prior probability measure on a Banach space, using random series through a Schauder basis. It is known that a Banach space with a Schauder basis is necessarily separable. Therefore, the following construction can practically only endow a separable Banach space with a probability measure. That means that we can always define a probability measure on a Banach space X, but the support of that probability, it is always contained in a separable subspace of X.

Definition 2.3.1. *A* **Schauder basis**, *for a Banach space X, is a sequence of elements $\phi_i \in X$ such that for every element $u \in X$ there exists a unique sequence of $u_i \in \mathbb{R}$ satisfying*

$$u = \sum_{i=1}^{\infty} u_i \phi_i$$

where the convergence of the above sequence is understood with respect to the norm of X.

Note that we understand the convergence in the above definition as follows: Let us consider partial sums,

$$u^n = \sum_{i=1}^{n} u_i \phi_i$$

then we have,

$$\|u^n - u\|_X \to 0.$$

The basic concept of constructing a prior measure on a Banach space X with a Schauder basis $\{\phi_i\}_{i=1}^{\infty}$ and element $m_0 \in X$, is based on the following random series,

$$u = m_0 + \sum_{i=1}^{\infty} u_i \phi_i \tag{2.7}$$

where $u_i = \gamma_i \xi_i$ with deterministic sequence $\gamma = (\gamma_i)_{i=1}^{\infty}$ and random sequence $\xi = (\xi_i)_{i=1}^{\infty}$ of independent and identical distributed ξ_i. Using the definition of the random element u, one can think of the prior μ_0 as the following measure: $\mathbb{P}(\xi \in \cdot)$ which defines a measure on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$, the push forward of this measure under $\xi \mapsto u$ gives the measure μ_0 on X.

Observe that both γ and ξ are used to define the appropriate probability over X, to keep it separate from $\{\phi_i\}$, we assume that $\|\phi_i\| = 1$. For fixed $\{\phi_i\}$ and ξ, the faster γ converges to zero, the smoother space X is. Notice also that if ξ is centred, i.e. the expectation of ξ_i is zero, the expected value of u is equal to m_0.

In the next section, we demonstrate two examples of priors defined on Banach spaces. Those have been highlighted in the literature.

2.3.1 Examples of priors

This section describes three examples of priors defined on Banach spaces based on the above construction. Those are uniform, Gaussian and Besov priors. For a more detailed proof, see for instance Dashti and Stuart (2015).

Uniform priors

In this example, we consider the Banach space $X = L^\infty(D)$ and we endow X with a uniform probability measure. Note that the specific space is not separable, but the above-mentioned procedure for the endowment of a Banach space with a probability measure requires a separable Banach space; at least, the support of that probability measure has to be contained in separable Banach space. For the definition of that separable Banach space, we take m_0 and a sequence of $\{\phi_i\}_{i=1}^\infty$ such that m_0, $\phi_i \in L^\infty(D)$, then the separable Banach space X' is defined as the span of that sequence with respect to the $\|\cdot\|_\infty$ norm, i.e. $(X', \|\cdot\|_\infty) = (m_0 + \text{span}\{\phi_i : i \in \mathbb{N}\}, \|\cdot\|_\infty)$.

Then, we can define the probability measure over the Banach space X as the push-forward probability measure of the following random series,

$$u = m_0 + \sum_{i=1}^{\infty} u_i \phi_i$$

where $u_i = \gamma_i \xi_i$ with deterministic sequence $\gamma \in l^1$, and random series ξ of independent and identical distributed with $\xi_i \sim U[-1, 1]$.

According to Theorem 2.1 in Dashti and Stuart (2015), one use appropriate condition on the sequence γ and m_0, which can control the push-forward measure emerging from the above random series, i.e. $\mu_0(\cdot) = \mathbb{P}(u \in \cdot)$, where u is defined through (2.7). For example, for given $a, b > 0$, one can choose γ and m_0 in such a way that the support of μ_0 is contained in $\{u : D \to \mathbb{R} \,|u(x) \in [a, b], \text{ for a.e.} x \in D\} \cap X'$. That can be especially useful in the case of the elliptic inverse problem, see for instance paragraph 1.3 in Dashti and Stuart (2015).

In the same notes and more precisely in Theorem 2.3, one can find also appropriate conditions on m_0, ϕ_i and γ, in order to define a uniform prior on Hölder space with β-exponent, i.e. $X = C^{0,\beta}(D)$.

Gaussian and Besov priors

This example builds the framework for priors over Besov and Sobolev spaces endowed with exponential probability measures. Let us start with the following Hilbert space

$$X = \dot{L}^2(\mathbb{T}^d) = \left\{ u : \mathbb{T}^d \to \mathbb{R} : \int_{\mathbb{T}^d} |u(x)|^2 dx < \infty, \int_{\mathbb{T}^d} u(x) dx = 0 \right\}$$

of real valued periodic functions $d \leq 3$ with inner-product $(\cdot, \cdot)_X$ and a norm $\|\cdot\|_X$, where $\mathbb{T}^d$ is the d-dimensional torus.

Consider $\{\phi_i\}_{i=1}^{\infty}$ to be an orthonormal basis of X, so that $u \in X$ is written as

$$u = \sum_{i=1}^{\infty} u_i \phi_i, \qquad \text{where } u_i = (u, \phi_i)_X$$

Then, we can define the Banach space $X^{t,q}$,

$$X^{t,q} := \left\{ u : \mathbb{T}^d \to \mathbb{R} : \|u(x)\|_{X^{t,q}} < \infty, \int_{\mathbb{T}^d} u(x)dx = 0 \right\}$$

where $\|u\|_{X^{t,q}} := \left(\sum_{i=1}^{\infty} i^{\left(\frac{tq}{d} + \frac{q}{2} - 1\right)} |u_j|^q \right)^{1/q}$ with $q \in [1, \infty)$ and $t > 0$. One can see that using a different basis, i.e. $\{\phi_i\}_{i=1}^{\infty}$, $X^{t,q}$ represent a different space, for example:

- If we choose $\{\phi_i\}_{i=1}^{\infty}$ to be the Fourier basis and $q = 2$, then the space $X^{t,2}$ represents $\dot{H}^t(\mathbb{T}^d)$, i.e. Sobolev space of periodic functions with mean-zero and square-integrable derivatives up to t.

- On the other hand, if we choose $\{\phi_i\}_{i=1}^{\infty}$ to be a wavelet basis, then the space $X^{t,q}$ represents the Besov space B^t_{qq}.

Then, using the random series which introduced in (2.7) with appropriate deterministic sequence γ and random sequence ξ, one can endow the space $X^{t,q}$ with a probability measure. In this example, we take ξ to be i.i.d with ξ_i centred q-exponential distribution, that is

$$\xi_i \sim c_q \mathrm{e}^{-|x|^q/2}$$

for some $q \in [1, \infty)$ and with c_q the normalising constant. For example for $q = 2$, we have that ξ_i are Gaussian distributed and for $q = 1$, ξ_i are Laplace-distributed. Also, for $s, \delta > 0$, we define the deterministic sequence γ as follows:

$$\gamma_j = j^{-\left(\frac{s}{d} + \frac{1}{2} - \frac{1}{q}\right)} \frac{1}{\delta^{1/q}}$$

According to Theorem 2.6 in Dashti and Stuart (2015), under the appropriate assumptions on t, s, q, we have that the support of the the push-forward measure emerging by (2.7) is contained in $X^{t,q}$.

An interesting example of $X^{t,q}$ is the 1-Besov priors with wavelets for $\{\phi_i\}_{i=1}^{\infty}$ and ξ are Laplace-distributed. According to Donoho and Johnstone (1998), B^t_{11} contains functions with considerable spatial inhomogeneity, i.e. it is extremely spiky in some parts of its domain and in other parts is extremely smooth. Also, there is an appropriate basis $\{\phi_i\}_{i=1}^{\infty}$ such that the most likely draws from the prior are described by relatively few non-zero coefficients u_i in basis $\{\phi_i\}_{i=1}^{\infty}$. These two properties are particularly interesting in the field of image processing and signal processing. In those fields, one needs to recover a solution which is rapidly changing its values, see for example ultrasound images. Those images are very smooth on some parts and it is changing rapidly on other parts.

Furthermore, using the same random series, we can endow Hölder spaces with a q-exponential probability measures. We only need to take a basis $\{\phi_i\}_{i=1}^{\infty}$ which consists of bounded and $\alpha-$Hölder functions ϕ_i, and both the bound of ϕ_i and its Lipschitz constant are growing with respect to i. Then, Theorem 2.8 in Dashti and Stuart (2015), implies that under appropriate conditions the support of the the push-forward measure emerging by (2.7), is contained in $C^{0,\beta}(\mathbb{T}^d)$.

2.4 Well-posedess of Bayesian approach

This section presents a set of conditions that ensure the well-posedness of the Bayesian approach. Thus, the following results are divided as follows: the posterior is well-defined, i.e. it satisfies existence and uniqueness, and stability under an appropriate metric, that is, small variation in the observed data causes small changes in the posterior distribution.

Let us consider X, Y to be separable Banach spaces equipped with the Borel σ-algebra, and μ_0 is a measure on X. We also assume that $\mu^y \ll \mu_0$ and the Radon-Nikodym derivative of μ^y with respect to μ_0 is given as follows

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z(y)} e^{-\Phi(u;y)}, \qquad \text{where } Z(y) = \int_X e^{-\Phi(u;y)} \mu_0(du) \tag{2.8}$$

Assumption 2.4.1. *Let $X' \subseteq X$ and assume that $\Phi \in C(X' \times Y; \mathbb{R})$. Assume further that there are functions $M_i : [0,\infty)^2 \to [0,\infty)$, $i = 1,2$ monotonic nondecreasing separately in each argument, and with M_2 strictly positive, such that for all $u \in X'$, $y, y_1, y_2 \in B_Y(0,r)$,*

$$-M_1(r, ||u||_X) \leq \Phi(u;y)$$

$$|\Phi(u;y_1) - \Phi(u;y_2)| \leq M_2(r, ||u||_X)\, \|y_1 - y_2\|_Y$$

Remark 2.4.1. *Note that in the case of an inverse problem with finite dimension data and with Gaussian noise, we always have that the corresponding $\Phi(u;y)$ satisfies the first inequality. For the second inequality, we need further to assume that the forward operator is Lipschitz, or even satisfies the condition of polynomial growth.*

In the following theorem, we see that, under appropriate assumptions, for any $y \in Y$, there exists a unique posterior distribution μ^y, the proof of this theorem can be found in Dashti and Stuart (2015).

Theorem 2.4.1. *Let Assumption 2.4.1 hold. Assume that $\mu_0(X') = 1$ and that $\mu_0(X' \cap B) > 0$ for some bounded set B in X. Assume additionally that, for every fixed $r > 0$,*

$$e^{M_1(r,||u||_X)} \in L^1_{\mu_0}(X;\mathbb{R}).$$

Then, for every $y \in Y$, $Z(y)$ given by (2.8) is positive and finite and the probability measure μ^y given by (2.8) is well-defined.

The following result is related to the stability of the posterior distribution μ^y which means that for y' in a small neighbourhood of y, i.e. for $0 < \delta \ll 1$, $y' \in B_{\|\cdot\|_Y}(y,\delta)$, the posterior $\mu^{y'}$ is close to μ^y. For the definition of this kind of stability, one needs to consider a metric on the probability space $\mathcal{M}(X) := \{$ probabilities π defined on $(X, \mathcal{B}(X))\}$. In this case, we consider the Hellinger distance,

$$d_{Hell}(\mu,\nu)^2 := \frac{1}{2}\int \left(\sqrt{\frac{d\mu}{d\zeta}} - \sqrt{\frac{d\nu}{d\zeta}}\right)^2 d\zeta, \qquad \text{where } \zeta \in M(X) \text{ with } \mu,\pi \ll \zeta$$

Then, we know that Hellinger distance defines the metric space $(M(X), d_{Hell})$. The following theorem provides the stability of the posterior distribution with respect to Hellinger distance and its proof can be found in Dashti and Stuart (2015).

Theorem 2.4.2. *Let Assumption 2.4.1 hold. Assume that $\mu_0(X') = 1$ and that $\mu_0(X' \cap B) > 0$ for some bounded set B in X. Assume additionally that, for every fixed $r > 0$,*

$$e^{M_1(r,||u||_X)}\left(1 + M_2(r,||u||_X)^2\right) \in L^1_{\mu_0}(X;\mathbb{R}). \tag{2.9}$$

Then there is $C = C(r) > 0$ such that, for all $y, y' \in B_Y(0,r)$

$$d_{Hell}\left(\mu^y, \mu^{y'}\right) \le C||y - y'||$$

2.5 Approximations of the posterior

This section demonstrates a continuity property of the posterior distribution based on Φ, see for example (2.5). The metric on the probability space $M(X)$ is considered to be the Hellinger distance.

In this case, we only need to focus on X, so we drop the subscript about the data y. Let us consider X Banach space and μ_0 prior defined on X. We consider μ and μ^N posterior distributions which are absolutely continuous with respect to μ_0, and are given by the following two formulas

$$\begin{aligned} \frac{d\mu}{d\mu_0}(u) &= \frac{1}{Z(y)}e^{-\Phi(u;y)}, \qquad \text{where } Z(y) = \int_X e^{-\Phi(u;y)}\mu_0(du) \\ \frac{d\mu^N}{d\mu_0}(u) &= \frac{1}{Z^N(y)}e^{-\Phi^N(u;y)}, \qquad \text{where } Z^N(y) = \int_X e^{-\Phi^N(u;y)}\mu_0(du) \end{aligned} \tag{2.10}$$

As we see the following assumption is analogous to the one in the former section.

Assumption 2.5.1. *Let $X' \subseteq X$ and assume that $\Phi \in C(X';\mathbb{R})$. Assume further that there are functions $M_i : [0,\infty) \to [0,\infty)$, $i = 1,2$ monotonic nondecreasing, and with M_2 strictly positive, such that for all $u \in X'$,*

$$-M_1(||u||_X) \leq \Phi(u)$$

$$-M_1(||u||_X) \leq \Phi^N(u)$$

$$|\Phi(u) - \Phi^N(u)| \leq M_2(||u||_X)\,\psi(N) \quad \textit{where } \psi(N) \to 0 \textit{ as } N \to \infty$$

Notice that the important result in the following theorem is that μ^N can be used as an approximation for the posterior distribution of μ, its proof can be found in Dashti and Stuart (2015).

Theorem 2.5.1. *Let Assumptions (2.5.1) hold. Assume that $\mu_0(X') = 1$ and that $\mu_0(X' \cap B) > 0$ for some bounded set B in X. Assume additionally that, for every fixed $r > 0$,*

$$e^{M_1(||u||_X)}\left(1 + M_2(||u||_X)^2\right) \in L^1_{\mu_0}(X;\mathbb{R}).$$

Then the probability measures μ and μ^N given by (2.10) are well-defined. Furthermore, there is a constant $C > 0$ such that, for all N sufficiently large, the following holds

$$d_{Hell}\left(\mu, \mu^N\right) \leq C\psi(N)$$

2.6 Monte Carlo approximation

In the previous framework, we start with a known prior probability and we are interested in approximating the posterior distribution. This approximation computes the posterior distribution indirectly. Essentially, we are interested in evaluating the so-called quantity of interest. Those quantities describe the statistical behaviour of the posterior distribution, for example they are: confidence intervals, mean values, variances, probabilities for certain events, etc. In most of the cases, those quantities are expected values of measurable functions with respect to the posterior distribution.

Consider a measurable function $\phi : D \to \mathbb{R}$, a probability space $(\Omega, \Sigma, \mathbb{P})$ and the random variable X which is distributed uniformly on the set D, i.e. $X \sim U(D)$. Suppose, we want to compute the integral

$$I(\phi, U(D)) = \int_D \phi(x)dx = \int_\Omega \phi(X(\omega))\mathbb{P}(d\omega)$$

One way for the approximation of those quantities is to use Monte Carlo simulations. This method arises from probability theory and more precisely an inspiration about this theory comes from the law of large number.

Let us consider the following method, for given $X_1, \dots X_N$ random variables independent and identically distributed by $X_1 \sim U(D)$, we define the approximation I_N as follows

$$I_N(\phi, U(D)) = \frac{1}{N} \sum_{n=1}^{N} \phi(X_n)$$

For simplicity, let us now denote $I(\phi, U(D))$, $I_N(\phi, U(D))$ by I, I_N. According to the Strong law of large number, if $X_1, X_2 \dots$ is a sequence of independent and identically distributed with finite mean and finite variance then we have that

$$I_N = \frac{1}{N} \sum_{n=1}^{N} \phi(X_n) \to \mathbb{E}\phi(X) = \int_D \phi(x) dx = I$$

Thus, based on this, we can build the basic Monte Carlo method. Another important factor for an approximation method is to estimate its convergence rate, or in other words, how fast our method converges. The answer to that question is given by the central limit theorem,

$$\sqrt{N}\,(I_N - I) \to N(0, \sigma^2), \qquad \text{in distribution}$$

where $\sigma^2 = Var(g(X))$. Notice that the central limit theorem does not depend on the dimension of X. Then, the convergence rate can be assessed and we get that the "error term" is of order $O(1/\sqrt{N})$ regardless of the dimension of X.

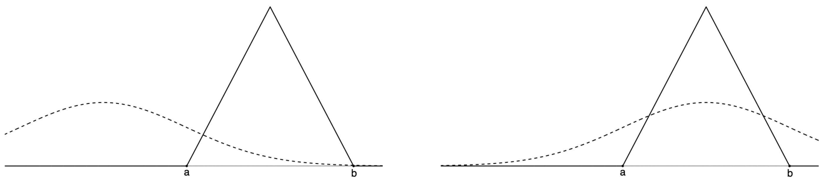
Let us consider the Riemann approximation and compare it with the above mentioned method. First, we assume $D = [0, 1]$ and define the Riemann approximation as follows,

$$\tilde{I}_N := \frac{1}{N} \sum_{n=1}^{N} \phi(y_n)$$

where $y_n = n/N$. For a smooth function, one can take that the convergence rate for this approximation is of order $O(1/N)$. Apparently, in this particular case the Riemann technique has better convergence rate in comparison to the Monte Carlo. On the other hand, deterministic approximations are not invariant to the change of the dimension. See for example that if we define the same technique on $D = [0, 1]^{10}$, in order to achieve the same level of accuracy $O(1/N)$, we need to evaluate $O(1/N^{10})$ points. Therefore, Monte Carlo simulation are more suitable for the computation of integrals in high-dimensional space.

The following example aims to present an idea for the acceleration of Monte Carlo based on a better sampling technique. Consider the function ϕ with support within the interval

$[a, b]$ and two probability measures μ, π with density functions f_μ, f_π, respectively. The following two Figures2.1.(a) and 2.1.(b) consists of the graphs of ϕ, f_μ and f_π. Notice that



(a) density of μ (b) density of π

Figure 2.1: The solid line refers to the function ϕ and the dashed line refers to the density function μ and π, respectively.

most of the mass of the probability μ is on the complement of the support of ϕ, while in case of π, it happens quite the contrary. Similar to the above mentioned Monte Carlo, we define the approximations $I_N(\phi, \pi)$ and $I_N(\phi, \mu)$ which approximate the integrals $I(\phi, \pi)$ and $I_N(\phi, \mu)$, respectively. We observe that most of X_i, in the case of $I_N(\phi, \mu)$, takes values on $[a, b]^c$, while in case of π, it happens quite the contrary. Therefore, we have that $I_N(\phi, \pi)$ approaches $I(\phi, \pi)$ faster than $I_N(\phi, \mu)$ approaches $I(\phi, \mu)$.

The above example suggests that the acceleration of the computation of a Monte Carlo method can be achieved using a sampling technique that concentrates most of the drawn points in a region, where ϕ takes its largest values.

In the next section, we study the importance sampling method. In that method, we consider two different measures: the first is the one we want to approximate, i.e. $I(\cdot, \mu)$, and the other is the one, we draw the sample from, i.e. $I_N(\cdot, \pi)$, those are the so-called target measure and proposal measure, respectively. In addition, we observe that the definition of that method provides a more suitable framework for the approximation of the posterior distribution.

2.6.1 Importance sampling

Let us consider a probability measure μ defined on $(X, \mathcal{B}(X))$ and a μ-integrable function $\phi : X \to \mathbb{R}$. The purpose of this method is the computation of the following integral using a similar probabilistic technique as the above relying on the (L.L.N.)

$$\mu(\phi) := \int_X \phi(x)\mu(dx)$$

In contrast with the above technique, the sample is not drawn directly from μ, but instead from an appropriate probability π, the so-called proposal probability. Also, in this point it is worth mentioning that μ is the so-called target probability measure.

The definition of the current method is based on the assumption that the Radon-Nikodym derivative of μ with respect to π, is known up to a multiplication constant

$$\frac{d\mu}{d\pi}(u) = g(u) \Big/ \int_X g(x)\pi(dx)$$

Since μ and π are measures, then $g : X \to \mathbb{R}_+$ is a non-negative function. Moreover, an equivalent condition for the existence of such g is that μ is absolutely continuous with respect to π. Notice that

$$\mu(\phi) := \int_X \phi(u)\frac{d\mu}{d\pi}(u)\pi(du) = \int_X \phi(u)g(u)\,\pi(du) \Big/ \int_X g(x)\pi(dx) =: \frac{\pi(\phi g)}{\pi(g)}$$

which is basically the inspiration for the following definition.

Definition 2.6.1. *The auto-normalised importance sampling estimator is given by:*

$$\begin{aligned} \mu^N(\phi) :=& \frac{\frac{1}{N}\sum\limits_{n=1}^{N} \phi(u^n)\, g(u^n)}{\frac{1}{N}\sum\limits_{m=1}^{N} g(u^m)} \qquad && u^n \sim \pi \ i.i.d \\ =& \sum_{n=1}^{N} w_n \phi(u^n) && where \quad w_n := \frac{g(u^n)}{\sum_{m=1}^{N} g(u^m)} \end{aligned}$$

Next, we examine the convergence of the auto-normalised importance sampling estimator. These results are presented as follows. First, we study the estimator's asymptotic consistency, and then we provide bounds for the bias of the estimator and the mean square error (MSE). Note that the last two quantities quantify the accuracy of the given estimator. On the other hand, asymptotic consistency tells us that the distribution of the estimator becomes more and more concentrated around the true value of $\mu(\phi)$ as the sample size becomes larger, i.e. $N \gg 1$.

The convergence of estimator $\mu^N(\phi)$ to $\mu(\phi)$ is a consequence of (L.L.N.) and the following property,

$$X_n \xrightarrow{\mathbb{P}} X, \quad Y_n \xrightarrow{\mathbb{P}} Y \qquad \Rightarrow (X_n, Y_n) \xrightarrow{\mathbb{P}} (X, Y)$$

Thus a sufficient condition for the convergence of the estimator is that both $\pi(g)$, $\pi(\phi g)$ should be finite. Similarly, for the application of the central limit theorem, we use Slutsky's theorem, which says that

$$X_n \xrightarrow{\text{in law}} X, \quad Y_n \xrightarrow{\mathbb{P}} c \qquad \Rightarrow X_n/Y_n \xrightarrow{\text{in law}} X/c$$

where c is constant. Therefore a sufficient condition for the application of the central limit theorem is that $\pi(g^2)$ and $\pi(\phi^2 g^2)$ are finite,

$$\sqrt{N}\left(\mu^N(\phi)-\mu(\phi)\right) \xrightarrow{\text{in law}} N\left(0, \frac{\pi\left(g^2\left(\phi-\mu(\phi)\right)^2\right)}{\pi(g)^2}\right)$$

The following estimations about bias and MSE are obtained in Agapiou et al. (2017). Therein, the author emphasises the importance of the quantity ρ, which is defined as the second moment of the Radon-Nikodym derivative of the target measure with respect to the proposal, i.e.

$$\rho = \frac{\pi(g^2)}{\pi(g)^2}$$

It is also easy to see, using Cauchy-Schwarz inequality, that $\rho \geq 1$. The proof of the following theorem can be found in Agapiou et al. (2017), see Theorem 2.1.

Theorem 2.6.1. *Assume that μ is absolutely continuous with respect to π, with square-integrable density g, that is, $\pi\left(g^2\right) < \infty$. The bias and MSE of importance sampling over bounded test functions may be characterized as follows:*

$$\sup_{|\phi|\leq 1}\left|\mathbf{E}\left[\mu^N(\phi)-\mu(\phi)\right]\right| \leq \frac{12\rho}{N}$$

and

$$\sup_{|\phi|\leq 1}\mathbf{E}\left[\mu^N(\phi)-\mu(\phi)\right]^2 \leq \frac{4\rho}{N}$$

It is easy to see that the above estimates do not apply only to the case where $|\phi| \leq 1$, but they can be used as a general statement for bounded functions. The extension for any bounded ϕ can be shown by using the estimator definition and $\mu(\phi)$. In addition, Theorem 2.3 in the same paper provides us with bounds on the bias and MSE for given ϕ which is not necessarily bounded.

We remind the example of the last section, see Figure 2.1, its purpose is to highlight that an appropriate proposal distribution can accelerate the computation of the Monte Carlo estimates. Let us re-state that example in a way that matches with the definition of the importance sampling. Suppose we have a given function ϕ and probability μ that admits a density function f_μ, then similarly to that example, we see that the best candidate for f_π is a function in the shape of the product of ϕ and f_μ.

Based on that example, we study the effective sample size (ess) which is defined as follows, see for instance Agapiou et al. (2017),

$$\text{ess}(N) := \left(\sum_{n=1}^{N}(w^n)^2\right)^{-1} = \frac{\left(\sum_{n=1}^{N} g\left(u^n\right)\right)^2}{\sum_{m=1}^{N} g\left(u^m\right)^2} = N\frac{\pi_{MC}^N(g)^2}{\pi_{MC}^N(g^2)}$$

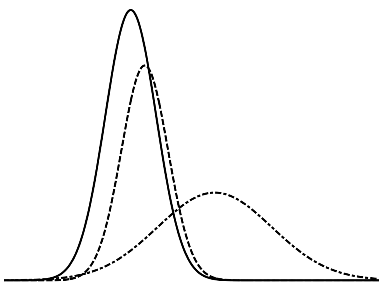
where π_{MC}^{N} is the empirical Monte Carlo estimate

$$\pi_{MC}^{N}(g) := \frac{1}{N}\sum_{n=1}^{N} g(u^{n}), \qquad u^{n} \sim \pi$$

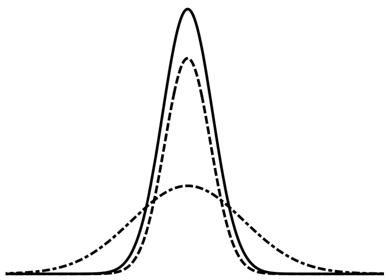
The following inequality provide the range of ess, the right-hand side of the inequality is a result of the Cauchy-Schwarz, for the the other side, w^n is less than or equal to 1

$$1 \leq \text{ess}(N) \leq N$$

In order to see that ess(N) quantifies the effectiveness of the sample, we show two extreme examples. Let us consider an N-sized sample $(u_1, u_2, \ldots, u_N)$. In the first case we assume that only one w^n is non-zero, then by definition ess$(N) = 1$. On the other hand, we consider a sample that all weights are equal, so we get ess$(N) = N$. Observe that between the former two examples, the most efficient sample is the second one. The last example is given through the following Figure 2.2, observe that for a typical N-sample drawn from π, it is more likely for the ess(N) of Figure 2.2(a) to be smaller than ess(N) of Figure 2.2(b).



(a) Low ess



(b) High ess

Figure 2.2: Notice that in both figures, there is a dashed line and a dot-dashed line, in order to spot the differences the dot-dashed line is chosen in both figure to be the lower function. The dashed line represent the density function of the target measure μ, the dot-dashed line the proposal measure π and the solid line represents the unnormalised Radon-Nikodym of $\mu \ll \pi$.

If we further assume that $\pi(g^2) < \infty$, then the Strong Law of the Large Number implies that

$$\frac{\pi_{MC}^{N}(g)^2}{\pi_{MC}^{N}(g^2)} \xrightarrow{a.s.} \frac{1}{\rho}$$

Or equivalently, we get that for sufficient large N, it holds that

$$\text{ess}(N) \approx \frac{N}{\rho}.$$

Therefore, for sufficient large N, the MSE can be controlled using ess(N)

$$\sup_{|\phi|\leq 1} \mathbf{E}\left[\mu^{N}(\phi) - \mu(\phi)\right]^2 \lessapprox \frac{4}{\text{ess}(N)}$$

2.6.1.1 High and infinite dimension case

The purpose of the following example is to highlight a particular behaviour of the Radon-Nikodym derivative, i.e. g, for probability measures μ, π defined in High-dimensional space.

Suppose two probability measures μ_1 and π_1 are defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with $g_1(u) = e^{-h(u)}$ and $h : \mathbb{R} \to \mathbb{R}_+$. In addition, we assume that the mean and the variance of μ_1 and π_1 are finite and also h is not constant in order to avoid the trivial case $\mu_1 = \pi_1$. Now, we consider the product probability measures μ_n and π_n on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ with

$$\mu_n(du) := \prod_{i=1}^{n} \mu_1(du_i), \qquad \pi_n := \prod_{i=1}^{n} \pi_1(du_i)$$

The unnormalised Radon-Nikodym is given by

$$g_n(u) = e^{-\sum_{i=1}^{n} h(u_i)}$$

Since h is non-negative, we have that has all the polynomial moments of g_n under π_n are finite. Moreover, we have that for every finite n, $\mu_n \ll \pi_n$, thus we can apply importance sampling on each one of them.

Let us define the set

$$A_{\mu_1} = \left\{ u \in \mathbb{R}^\infty : \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} u_i = \int x \mu_1(dx) \right\}$$

and μ_∞ and π_∞ are the limit probability measures of μ_n and π_n, respectively.

Next, we want to evaluate $\mu_\infty(A_{\mu_1})$ and $\pi_\infty(A_{\mu_1})$. Notice also that

$$\int x\mu_1(dx) = \int x \frac{g_1(x)}{\pi_1(g_1)} \pi_1(dx) \neq \int x\pi_1(dx)$$

Using the Strong Law of Large Number, one can prove that

$$\mu_\infty(A_{\mu_1}) = 1, \qquad \pi_\infty(A_{\mu_1}) = 0$$

Thus, μ_∞ and π_∞ are mutually singular. The above mentioned example taught as that any two measure μ_∞ and π_∞ which differ, in the above sense, in infinite dimension without that difference is vanishing as long n goes to infinity, then we will always have that those two measures, μ_∞, π_∞, are mutually singular. Note that since μ_∞ and π_∞ are mutually singular, thus there is no Radon-Nikodym derivative and hence importance sampling can not be defined.

Based on the singularity, it is worth investigating how this affects MSE as the dimension of the problem increases. Remember that ρ/N controls MSE. Using that h is not constant, we get that $\rho_1 > 1$, and also due to the structure of μ_n and π_n, we have $\rho_n = (\rho_1)^n$. Let us say we are interested in achieving the same accuracy, i.e. MSE, as n goes to infinity. In order to compensate the exponential growth of ρ_n, we need to increase the sample size exponentially with respect to n.

Essentially, the above example says that if each coordinate of g plays a significant role for the construction of μ_∞, then the probability measures μ_∞ and π_∞ are mutually singular. As a consequence, we have that for non-trivial probabilities μ and π defined on the high dimensional space X with $\mu \ll \pi$, the part of X that contributes most to the evaluation, tends to be limited to a low-dimensional subspace.

Motivated by these considerations, we are interested in a distance that can quantify the significance of each coordinate of g. In order to determine that distance, we use the exponential structure of ρ that emerges from the above example. Therefore, we introduce the so-called Kullback-Leibler divergence.

Let us assume two probability measures μ and π are defined over the same measurable space $(X, \mathcal{B}(X))$. Then their Kullback-Leibler divergence is defined as follows

$$D_{KL}(\mu||\pi) = \int \log\left(\frac{d\mu}{d\pi}\right) d\mu \tag{2.11}$$

Notice that

$$D_{KL}(\mu||\pi) = \begin{cases} < \infty, & \mu \ll \pi \\ \infty, & \text{otherwise} \end{cases}$$

Thus it is obvious that Kullback-Leibler is not symmetric.

Using Jensen inequality, one can show that

$$D_{KL}(\mu||\pi) \geq \log\left(\int \frac{d\mu}{d\pi} d\mu\right) = 0 \tag{2.12}$$

In addition, it may be shown that

$$e^{D_{KL}(\mu||\pi)} \leq \rho$$

Let us recall that ρ/N controls the MSE, see Theorem 2.6.1, thus we have that N has to be at least exponentially larger than $D_{KL}(\mu||\pi)$.

Chapter 3

Stochastic differential equations

In this chapter we discuss stochastic differential equations. We review the theory of existence of solutions and well-posedness and their corresponding forward and backward Kolmogorov equations. As we shall be considering these equations on bounded intervals later in the thesis, we shall also look at how the process is defined in a neighbourhood of the boundary points.

3.1 Existence and uniqueness in $\mathbb{R}$

We consider a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t\geq 0}$ that is right-continuous and $\mathcal{F}_0$ contains all $\mathbb{P}-$null sets. Also, $(W_t)_{t\geq 0}$ is a brownian motion defined on that space. Let consider $\mu : [0,T] \times \mathbb{R} \to \mathbb{R}$ and $\sigma : [0,T] \times \mathbb{R} \to \mathbb{R}$ to be measurable functions. Then, we can define a stochastic differential equation of Itô type with initial values, as follows

$$\begin{aligned} &dX_t = \mu\left(t, X_t\right) dt + \sigma\left(t, X_t\right) dW_t,\ t \in [0,T] \\ &X(t_0) = x_0 \end{aligned} \tag{3.1}$$

There are several frameworks that one consider in oder to define a solution to the above equation. Let start with the most common structure for a solution that satisfies (3.1) and summarise the structure in the following definition, this definition comes from Evans (2012) book.

Definition 3.1.1. *A real-valued stochastic process $(X(t))_{t\in[0,T]}$ is called solution of (3.1) with initial value $X(t_0) = x_0$, if the following hold:*

(i) $(X(t))_{t\in[0,T]}$ *is continuous for a.e. path and $\mathcal{F}_t$ adapted*

(ii) $\mathbb{E}\int_0^T |\mu\left(s, X_s\right)| ds < \infty$ *and* $\mathbb{E}\int_0^T |\sigma\left(s, X_s\right)|^2 ds < \infty$

(iii) equation (3.1) holds for every $t \in [0,T]$ with probability 1.

The question that naturally comes to a person who is interested in the solution of such equations is, whether there are more than one solution to this problem. There are several definitions about the uniqueness of such problems, but specifically in this framework the one we consider is the so-called path-wise uniqueness and it is given in the following terms:

We say that the solution of equation (3.1) is unique, if for any processes $(X(t))_{t\in[0,T]}$ and $\left(\overline{X}(t)\right)_{t\in[0,T]}$ which solves equation (3.1) holds that

$$\mathbb{P}\left(X(t) = \overline{X}(t),\, \forall t \in [0,T]\right) = 1$$

The following theorem gives the most common assumptions under which equation (3.1) has a unique solution, for more details about its proof see for instance Theorem 3.1 in Mao (2007),

Theorem 3.1.1. *Assume functions μ and σ satisfy the following two properties:*

(i) (Lipschitz condition) there exists a constant C such that: $\forall\, x, y \in \mathbb{R}$ and $t \in [0,T]$

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq C|x-y|$$

(ii) (Linear growth condition) there exists a constant C' such that: $\forall\, x \in \mathbb{R}$ and $t \in [0,T]$

$$|\mu(t,x)| + |\sigma(t,x)| \leq C'\,(1+|x|)$$

Then, there is a unique solution X for equation (3.1).

For the purpose of the present thesis, we study the particular case,

$$\begin{aligned} dX_t &= \mu\left(X_t\right)dt + \sqrt{\sigma\left(X_t\right)}dW_t,\ t \in [0,T] \\ X_{t_0} &= x_0 \end{aligned} \tag{3.2}$$

where μ, σ are Lipschitz and $\sigma \geq 0$. As, we can see the so-called diffusion coefficients of (3.2) is 1/2-Holder, so it does not fulfill the requirements of theorem 3.1.1. We can easily see that $\sqrt{\sigma\left(\cdot\right)}$ is 1/2-Holder because it satisfies the following inequality,

$$|\sqrt{\sigma(x)} - \sqrt{\sigma(y)}| \leq \sqrt{|\sigma(x)-\sigma(y)|} \leq \sqrt{C}\sqrt{|x-y|} \tag{3.3}$$

The following theorem will provide us with the existence of solutions to (3.2). Before that theorem, one needs to extend the definition provided above for the solution of equation (3.2). The following definition determines a weak solution, see for instance Karatzas and Shreve (2014).

Definition 3.1.2. *A weak solution of (3.1) with initial condition x is a stochastic process $(X(t))_{t\geq 0}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that for some Brownian motion $(W(t))_{t\geq 0}$ and some filtration $(\mathcal{F}_t)_{t\geq 0}$ with the usual condition[1], X satisfies the following:*

[1] A filtration $\{\mathcal{F}_t\}_{t\geq 0}$ is said to satisfy the usual conditions, if it is right-continuous and $\mathcal{F}_0$ contains all the $\mathbb{P}$-negligible events in $\mathcal{F}$

(i) $(X(t))_{t\geq 0}$ *is continuous a.e. and* $\mathcal{F}_t$ *adapted*

(ii) $\omega -$ *a.e., we have that:* $\int_0^T |\mu(s, X_s)| + \int_0^T |\sigma(s, X_s)|^2 ds < \infty$ *for every* $T > 0$,

(iii) equation (3.1) *holds for every* $t \geq 0$ *with probability 1.*

We see that the following theorem relaxes the condition for the existence of such a solution but also refers to a more general solution, for the proof of the following theorem, see page 60 of Skorokhod (1982).

Theorem 3.1.2. *Consider equation* (3.1) *with coefficients* μ, $\sigma \in C([0,T]\times\mathbb{R})$ *and assume that* μ *and* σ *satisfy the linear growth condition from Theorem 3.1.1. Then, equation* (3.1) *has at least a solution which is bounded with probability* 1.

Using inequality (3.3) and the Lipschitz continuity of μ, we can obtain a constant $C > 0$ in the following way

$$|\mu(x)| + |\sqrt{\sigma(x)}| \leq \mu(0) + C_\mu|x| + \sigma(0) + \sqrt{C_\sigma}\sqrt{|x|} \leq C(1 + |x|)$$

Hence, we can see that the coefficients of equation (3.2) satisfy Theorem 3.1.3.

Let us now state the following theorem which provide us the uniqueness for the solution of equation (3.2), the proof of that Theorem can be found in Yamada and Watanabe (1971).

Theorem 3.1.3. *Consider the equation,*

$$dX = \mu(X_t)dt + \sigma(X_t)dW_t. \tag{3.4}$$

Suppose that

(i) there exists a positive increasing function $\rho(x)$ *for* $x \in (0,\infty)$ *such that:*

$$|\sigma(x) - \sigma(y)| \leq \rho(|x - y|), \qquad \forall\, x, y \in \mathbb{R}$$
$$\text{and} \int_{0^+} \rho^{-2}(s)ds = +\infty$$

(ii) there exists a positive increasing concave function $\kappa(x)$ *for* $x \in (0,\infty)$ *such that:*

$$|\mu(x) - \mu(y)| \leq \kappa(|x - y|), \qquad \forall\, x, y \in \mathbb{R}$$
$$\text{and} \int_{0^+} \kappa^{-1}(s)ds = +\infty.$$

Then, the solution is pathwise unique.

3.2 SDEs on bounded sets

In this section, we consider processes which take values in an I sub-interval of the real line. In particular, we have that the equation in (3.1) determines the behaviour of the process only in the interior of interval I, where for the boundary, we need to take further assumptions.

According to Feller (1954a) and Feller (1954b), in one dimension, there are only four kinds of behaviours that a process can have on the boundary in order for this to have continuous paths:

(i) absorption: once the process reaches the boundary, it remains there.

(ii) instantaneous reflection: when the process reaches the boundary, it returns to the interior of the interval in a continuous way; moreover, the time that almost each path spends on the boundary is zero.

(iii) delayed reflection: similarly as the above, but instead of leaving instantaneously, it remains for a positive time, so for almost each path, we have that the time the process spends on the boundary is positive.

(iv) partial reflection: this is a combination of absorption and instantaneous reflection, which means that once the process reaches the boundary it either remains there, or reflects instantaneously.

In the case of an SDE with absorption on the boundary, this is the so-called SDE with absorbing barrier, Doob (1955) suggests that the solution of an SDE with absorbing barrier can be realised as the stopped process of the solution of the equation (3.1). In particular, the author defines the standard Itô process as the solution X_t of equation (3.1) under the conditions that the coefficients: μ and σ are Baire functions with respect to the pair (t, x), also these are satisfying the Lipschitz and the linear growth conditions, see (3.1.1), and last σ is positive. Then, if X_t is a standard Itô process and τ is a stopping time, the process $X_{t\wedge\tau}$ is given by the following equation

$$X\left(t\wedge\tau\right)-X\left(0\right)=\int_0^t\mu\left(s,X_s\right)\,1\left(\tau>s\right)\,ds+\int_0^t\sigma\left(s,X_s\right)\,1\left(\tau>s\right)\,dW_s$$

Also, in the same paper, one can find the following theorem which provides the uniqueness conditions.

Theorem 3.2.1. *Consider two standard Itô processes* $(X_1(t))_{t\in[0,T]}$ *and* $(X_2(t))_{t\in[0,T]}$ *with initial value* $X_i(0) = x$ *and coefficients* μ_i *and* σ_i*, for* $i = 1, 2$*, respectively. Assume that* $X_i(t)$*,* $i = 1, 2$ *are defined with the same Brownian motion* $W(t)$*. Consider a closed interval* I *(finite, or infinite) such that for every* $t \in [0, T]$ *and* $y \in I$*,* μ_i *coefficients coincide and similarly for* σ_i*, and, define the stopping times* $\tau_i = \inf\{t : X_i(t) \in \partial I\}$*. Then, if* $x \in I$*, we have that:*

$$\mathbb{P}\left(\tau_1=\tau_2\right)=1,\qquad \mathbb{P}\left(X_1(t)=X_2(t),\,\forall t\leq\tau_1\right)=1$$

For boundary cases (ii)-(iv) above, one can find the conditions for existence and uniqueness of the process in Skorokhod (1961) and Skorokhod (1962). The following is an example of a reflecting SDE with instantaneous reflection on the boundary, the so-called SDE with reflecting barrier. Let us consider the case, where the domain of the process is the half-line $\{x : x \geq 0\}$, and the boundary point is 0. The above references suggest that the solution should be sought as a pair of two processes (ξ, l) which satisfy the following equation

$$\xi\left(t\right) - \xi\left(0\right) = \int_0^t \mu\left(s, \xi_s\right) ds + \int_0^t \sigma\left(s, \xi_s\right) dW_s + l(s) \tag{3.5}$$

where l satisfy the following properties: for almost every path, l is a continuous monotone function and the points of growth can only occur when the process ξ attains the boundary, i.e. $\xi = 0$, and the set $\{t : \xi(t) = 0\}$ has Lebesgue measure 0. Also in Skorokhod (1962), we can find a different representation for the process l from the one that follows in the next section.

In the following two subsections, we focus on the above example in order to develop some necessary tools that we need later for this thesis. In the first subsection, we introduce the Skorokhod problem and the corresponding Skorokhod map, while in the second subsection, we use that map in order to obtain an equivalent differential form of equation (3.5), where that new differential form looks similar to (3.1).

Our main interest is to focus on these two cases, so we are not going to provide any further details about boundaries as in case (*iii*), or (*iv*).

3.2.1 Skorokhod problem

In this problem, we consider a continuous function f and we are interested to divide it in a unique way into the difference of two positive functions g and l. The following is a more rigorous definition of that problem, see for instance in Pilipenko (2014).

Definition 3.2.1. *For a given $f \in C\left([0,T]\right)$ with $f(0) \geq 0$, a pair of continuous functions g and l are called a solution of the Skorokhod problem for f if*

(i) $g(t) \geq 0, \ t \in [0,T]$

(ii) $l(0) = 0$ *and l is non-decreasing w.r.to t*

(iii) $\int_0^T 1\left(g(s) > 0\right) dl(s) = 0$

(iv) $g(t) = f(t) + l(t), \qquad \forall t \in [0,T]$

The following theorem provides the existence and uniqueness of the Skorokhod problem's solution; the proof is omitted but it can be found in Skorokhod (1962).

Theorem 3.2.2. *Suppose $f \in C([0,T])$ and $f(0) \geq 0$. Then, the Skorokhod problem has a unique solution. Furthermore, the solution is given as it follows:*

$$l(t) = -\min_{s\in[0,t]} \{f(s) \wedge 0\} = \max_{s\in[0,t]} \{(-f(s)) \vee 0\}$$
$$g(t) = f(t) + l(t) = f(t) - \min_{s\in[0,t]} \{f(s) \wedge 0\}$$

Before, we return in the case of the reflecting barrier, we define Skorokhod map as it follows:

$$g(\cdot) = \Gamma f(\cdot) = f(\cdot) - \min_{s\in[0,\cdot]} \{f(s) \wedge 0\} \tag{3.6}$$

The following Lemma summarises some of the Skorokhod map properties, the proof of which is omitted, but it is easy to obtain and can be found in Pilipenko (2014).

Lemma 3.2.1. *Let consider the Skorokhod map $\Gamma : (C([0,T]), \|\cdot\|) \to (C([0,T]), \|\cdot\|)$, where $\|f\|_{[0,t]} := \sup_{s\in[0,t]} |f(s)|$. Then, the following holds:*

(i) for every $f_1, f_2 \in C([0,T])$ and $t \in [0,T]$

$$\|g_1 - g_2\|_{[0,t]} \leq 2\|f_1 - f_2\|_{[0,t]}$$
$$\|l_1 - l_2\|_{[0,t]} \leq \|f_1 - f_2\|_{[0,t]}$$

where $g_i = \Gamma f_i$ and $l_i = f_i - g_i$ for $i = 1, 2$

(ii) for every $\delta > 0$ and $f \in C([0,T])$:

$$\omega_g(\delta) \leq \omega_f(\delta), \qquad \omega_l(\delta) \leq \omega_f(\delta)$$

where $g = \Gamma f$, $l = f - g$ and $\omega_f(\delta) := \sup_{\substack{t,s\in[0,T] \\ |t-s|<\delta}} |f(t) - f(s)|$

(iii) for every $f \in C([0,T])$ and $t \in [0,T]$

$$\|\Gamma f\|_{[0,t]} \leq 2\|f\|_{[0,t]}, \qquad \|l\|_{[0,t]} \leq \|f\|_{[0,t]}$$

From point (i) of the above Lemma, it is immediate to conclude that Skorokhod map is continuous.

3.2.2 Existence and uniqueness of SDE's solution with reflecting barrier

Let us consider once again the example that introduced in (3.5),

$$\begin{gathered} d\xi_t = \mu(t,\xi_t)dt + \sigma(t,\xi_t)dW_t + dl_t, \qquad t \geq 0 \\ \text{with reflection at 0 and initial condition } \xi(0) = \xi_0. \end{gathered} \tag{3.7}$$

The following definition summarises all the properties suggested in Skorokhod (1961) and can be found in Pilipenko (2014).

Definition 3.2.2. *A pair of continuous $\mathcal{F}_t$-adapted processes $(\xi_t, l_t)_{t\geq 0}$ is called solution of the stochastic differential equation (3.7), if $\xi(t)$ and $l(t)$ satisfy the following:*

(i) $\xi(t) \geq 0$ *for* $t \geq 0$

(ii) l *is non-decreasing and* $l(0) = 0$

(iii) $\int_0^t 1\,(\xi(s) > 0)\,dl(s) = 0, \;\; t \geq 0$

(iv) For almost every ω, we have that: for $t \geq 0$ it holds that

$$\xi(t) = \xi_0 + \int_0^t \mu(s, \xi(s))ds + \int_0^t \sigma(s, \xi(s))dW_s + l(t) \tag{3.8}$$

and all integrals are well-defined.

We now show that equation (3.7) and the Skorokhod problem can be combined in such a way that we can obtain an equivalent differential form to equation (3.8).

Let us assume that there exists a pair of processes following definition 3.2.2 and fix ω that satisfies equation (3.8), and define

$$Y(t) = \xi_0 + \int_0^t \mu(s, \xi(s))ds + \int_0^t \sigma(s, \xi(s))dW_s, \qquad t \geq 0. \tag{3.9}$$

According to the definition 3.2.2, the integrals are well-defined which implies that the process Y is continuous. Observe that for fixed ω, we can apply Theorem 3.2.2. We only need to see that $\xi(t, \omega)$ satisfies the same assumption as the one considered for the function $g(t)$ in definition 3.2.1, and also the same holds for $l(t, \omega)$ and the corresponding $l(t)$ comes from the definition 3.2.1. The uniqueness of Theorem 3.2.2 provides us that there is a unique pair $(\xi(t, \omega), l(t, \omega))$ which satisfies the definition in 3.2.1 and defines $Y(t, \omega)$. Therefore, using the representation for the solution of the Skorokhod problem, we get that

$$\xi(t, \omega) = \Gamma Y(t, \omega), \;\; t \geq 0$$

Hence, equation (3.9) can be written as follows:

$$Y(t) = \xi_0 + \int_0^t \mu(s, \Gamma Y(s))ds + \int_0^t \sigma(s, \Gamma Y(s))dW_s, \;\; t \geq 0 \tag{3.10}$$

Conversely, assuming a process Y that satisfies equation (3.10). The application of Theorem 3.2.2 on Y provides us with such ξ and l which satisfy definition 3.2.2.

Since equations (3.8) and (3.10) are equivalent, it make sense to take the following notation: let consider equation (3.8) with coefficients $\mu(t, \cdot)$ and $\sigma(t, \cdot)$, then the coefficients of equation (3.10) can be denoted by $\tilde{\mu}(t, \cdot) := \mu(t, \Gamma(\cdot)_t)$ and $\tilde{\sigma}(t, \cdot) := \sigma(t, \Gamma(\cdot)_t)$. By using

Lemma 3.2.1, it is easy to see that if μ and σ satisfy Lipschitz continuity and linear growth, as in Theorem 3.1.1, then $\tilde{\mu}$ and $\tilde{\sigma}$ satisfy similar properties:

$$\begin{aligned} |\tilde{\mu}(t, f_1) - \tilde{\mu}(t, f_2)| + |\tilde{\sigma}(t, f_1) - \tilde{\sigma}(t, f_2)| &\leq 2C\|f_1 - f_2\|_{[0,t]} \\ |\tilde{\mu}(t, f)| + |\tilde{\sigma}(t, f)| &\leq 2C'\left(1 + \|f\|_{[0,t]}\right) \end{aligned} \qquad \forall f, f_1, f_2 \in C([0,T]) \text{ and } t \in [0,T]$$

where the above $\|\cdot\|$ is the supremum-norm, the same as the one, we use in the last section. The following theorem is the analogue of Theorem 3.1.1, the proof of which is omitted, but we can find it in two different forms, either by using the above observation for the coefficients of equation (3.10) and then apply the proof of Theorem 2.2 on page 150 in Mao (2007), or we can use an alternative proof based on equation (3.8), see Theorem 1.2.1 in Pilipenko (2014).

Theorem 3.2.3. *Let ξ_0 be a non-negative $\mathcal{F}_0$-adapted random variable and also assume functions μ and σ satisfying the following two properties:*

(i) (Lipschitz condition) there exists constant C such that: $\forall\, x, y \in \mathbb{R}$ and $t \in [0,T]$

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq C|x-y|$$

(ii) (Linear growth condition) there exists constant C' such that: $\forall\, x \in \mathbb{R}$ and $t \in [0,T]$

$$|\mu(t,x)| + |\sigma(t,x)| \leq C'\left(1 + |x|\right)$$

Then, there exists a unique solution according to the definition 3.2.2.

Similarly, as in section 3.1, we are also interested to study the analogous equation to equation (3.2), i.e. the case where μ and σ are Lipschitz, $\sigma \geq 0$ and ξ, l satisfy the following equation:

$$\begin{aligned} &d\xi_t = \mu\left(\xi_t\right) dt + \sqrt{\sigma\left(\xi_t\right)} dW_t + dl_t,\ t \in [0,T] \\ &\xi(0) = \xi_0 \end{aligned} \tag{3.11}$$

By using the same notation for $\tilde{\mu}$ and $\tilde{\sigma}$, as the one is used earlier in this section, applying part (iii) of Lemma 3.2.1 and the Lipschitz property of μ and σ, we are able to obtain that $\tilde{\mu}$ and $\sqrt{\tilde{\sigma}}$ satisfy the linear growth condition:

$$|\mu(\Gamma(Y)_t)| + |\sqrt{\sigma(\Gamma(Y)_t)}| \leq |\mu(0)| + |\sqrt{\sigma(0)}| + 2C\left(|\Gamma(Y)_t| + |\Gamma(Y)_t|^{1/2}\right) \leq \tilde{C}(1 + \|Y\|_{[0,t]})$$

The following theorem provides us the uniqueness of the solution of equation (3.11) and its proof can be found in section 3.4.1 below. Therein, we combine the following two proofs: Theorems 1 from Yamada and Watanabe (1971) and Theorem 1.2.1 from Pilipenko (2014).

Theorem 3.2.4. *Let ξ_0 be a non-negative $\mathcal{F}_0$-adapted random variable and μ and σ from equation (3.11) satisfy the Lipschitz condition. Then, Equation 3.11 has a unique solution.*

3.3 Fokker-Planck equation

The purpose of the current section is to introduce Fokker-Planck equations and associate them with the solution of parabolic and elliptic equation. We open this sections with the demonstration of the generic form of Kolmogorov equations, backward and forward. Then, an introduction to semigroups defined over a Markov process follows and then, we provide several examples of semigroups which represent the solution of parabolic and elliptic equations.

Let X be the strong solution of equation (3.1), which also admits a transition probability density function $p(t,y;s,x)$, i.e.

$$\mathbb{P}\left(X_t \in dy | X_s = x\right) := p(t,y;s,x)\,dy, \qquad \text{for } s<t$$

It is worth to point out that the transition probability density function p has two set of variables (t,y) and (s,x) which represent the current position (at time s) and the future position, respectively. Therefore, a Kolmogorov equations is called backward, or forward respectively, based on which set of variables is differentiated. For a better visualisation of the two differential forms, we choose to omit the set of variables that does not contribute to the problem. A backward Kolmogorov equation is given in the following terms

$$\begin{gathered} -\frac{\partial\, p(s,x)}{\partial s} = \frac{1}{2}\sigma^2(s,x)\,\frac{\partial^2\, p(s,x)}{\partial x^2} + \mu(s,x)\,\frac{\partial\, p(s,x)}{\partial x} \\ \lim_{s \nearrow t} p(t,y;s,x) = \delta(y-x) \end{gathered} \tag{3.12}$$

A forward Kolmogorov equation, which also called Fokker-Planck equation, is given in the following terms

$$\begin{gathered} \frac{\partial\, p(t,y)}{\partial t} = \frac{1}{2}\,\frac{\partial^2}{\partial y^2}\left(\sigma^2(t,y)p(t,y)\,\right) - \frac{\partial}{\partial y}\left(\mu(t,y)p(t,y)\,\right) \\ \lim_{t \searrow s} p(t,y;s,x) = \delta(y-x) \end{gathered} \tag{3.13}$$

Next, an example follows which aims to highlight the relation between solution of equation (3.1) and the backward Kolmogorov equation. Consider a differential operator in the form of $L(\cdot) = \mu(x)\frac{d}{dx}(\cdot) + \frac{1}{2}\sigma^2(x)\frac{d^2}{dx^2}(\cdot)$, a continuous function f which is non-negative, or it has a polynomial growth, and also the following equation

$$\begin{aligned} -\frac{\partial u}{\partial t} &= Lu, \qquad \text{for } s \in [0,t),\ x \in \mathbb{R} \\ u(t,x) &= f(x) \end{aligned}$$

Let p satisfy backward Kolmogorov equation. Next, we define u to be given by the following expression, essentially u is defined through the following semigroup which act over the function f,

$$u(s,x) = \int f(y)p(t,y;s,x)dy$$

Now, we take the derivative of u with respect to t and see that

$$\begin{aligned}
-u_s(s,x) &= \int f(y)\left(-\frac{\partial p(t,y;s,x)}{\partial s}\right)dy \\
&= \int f(y)\left(\frac{1}{2}\sigma^2(s,x)\frac{\partial^2}{\partial x^2}p(t,y;s,x) + \mu(s,x)\frac{\partial}{\partial x}p(t,y;s,x)\right)dy \\
&= Lu(s,x)
\end{aligned}$$

Notice that $\lim_{s\nearrow t} p(t,y;s,x) = \delta(y-x)$ implies that

$$\lim_{s\nearrow t} u(s,x) = f(x)$$

The formula that relates the solution of equation (3.1) and of a parabolic equation is known as Feynman–Kac formula, for its proof see for instance the proof of Theorem 7.6 in Karatzas and Shreve (2014).

Next, a brief introduction of semigroups defined over homogeneous Markov process is presented, that builds a concept which can be applied on solutions of a homogeneous SDE, see for example (3.4). This concept can also be extended in order to cover the solutions of SDEs in the generic form of (3.1).

3.3.1 Markov process and semigroups

Let us consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Markov Process X_t which takes values on X, where $(X, \mathcal{B}(X))$ is a measurable space, with transition probability

$$P(t,\Gamma;s,x) := \mathbb{P}(X_t \in \Gamma | X_s = x), \qquad \text{for } 0 < s < t,\ \Gamma \in \mathcal{B}(X)$$

We consider that the transition probability is stationary, that means

$$P(t,\Gamma;s,x) =: P(t-s,x,\Gamma).$$

We also say that a transition probability is stochastically continuous, if for any $x \in X$ and any neighbourhood Γ contains x holds,

$$P(t,x,\Gamma) \to 1, \qquad t \downarrow 0.$$

Also, Markov property provides us with the so-called Chapman-Kolmogorov identity,

$$P(t+s,x,\Gamma) = \int_X P(s,y,\Gamma)P(t,x,dy).$$

Let us denote B to be the Banach space of bounded measurable functions on X endowed with the supremum-norm $\|f\| = \sup_{x\in X} |f(x)|$. Now, we can associate a Markov process with a family of operators T_t acting over the space B, for $t \geq 0$, in the following way,

$$T_t f(x) = \mathbb{E}_x f(X_t) = \int_X f(y) P(t, x, dy), \qquad \forall f \in B.$$

As we can see, T_t is a contraction linear map[2] and satisfies the semigroup property[3], the latter being a direct consequence of Chapman-Kolmogorov identity. Similarly, we denote V to be the Banach space of finite and finitely additive measures on $\mathcal{B}(X)$, the norm of V is the total variation, and then, we can define the semigroup U_t acting over the space V, as follows

$$(U_t\mu)(\Gamma) = \int_X P(t, x, \Gamma)\mu(dx), \qquad \forall \mu \in V, \Gamma \in \mathcal{B}(X)$$

As we can see V is the dual space of B, so the following equality comes to highlight that T_t and U_t are conjugate operators

$$\int_X T_t f(x)\mu(dx) = \int_X f(x)(U_t\mu)(dx), \qquad f \in B, \mu \in V$$

Let us now focus on T_t semigroup and define its infinitesimal generator A

$$Af := \lim_{t\downarrow 0} \tfrac{1}{t}\left(T_t f - f\right)$$

where the limit is with respect to the norm of B, i.e. $\lim_{t\downarrow 0} \| \frac{1}{t}(T_t f - f) - Af\| = 0$. As, we observe in order to define Af, we need at least to be restricted on those f which satisfy $\lim_{t\downarrow 0} ||T_t f - f|| = 0$

$$B_0 := \left\{ f \in B : \lim_{t\downarrow 0} ||T_t f - f|| = 0 \right\}$$

Let us now denote with $D(A)$ the domain of the infinitesimal generator A. Then, we know the following about it: $D(A)$ is a vector space and it is everywhere dense in the space B_0, see for instance Theorem 1.4 in Dynkin (1965). In the same book Theorem 2.3 states that every stochastically continuous P is uniquely determined by its infinitesimal A. Also, we have that for every for every $f \in D(A)$, the corresponding Cauchy problem,

$$\begin{aligned} &\frac{\partial u_t(x)}{\partial t} = Au_t(x), \qquad \text{where } u_t(x) = u(t,x) \\ &\lim_{t\downarrow 0} u(t,x) = f(x) \end{aligned} \tag{3.14}$$

has a unique solution within the class of bounded functions: $u(t,x) = T_t f(x)$, see Theorem 1.3 in Dynkin (1965).

[2] That means that T_t satisfy the inequality $\|T_t f\| \leq \|f\|$

[3] A mapping $T_{(\cdot)} : \mathbb{R}_+ \to L(X;X)$ is called semigroup, if satisfy the following two properties: $T_0 = I$, $T_{t+s} = T_t \cdot T_s$, where the product of two semi-groups is given by the composition of those two mappings.

Similarly, we can define the infinitesimal generator A^* for the semigroup U_t, for a $\mu \in D(A^*)$, we can define a function $\nu_t(\Gamma) = U_t\mu(\Gamma)$ which is the solution of the corresponding Cauchy problem.

3.3.2 Association between diffusion processes and differential equations

In diffusion theory one usually considers that the transition probability $P(t, x, \Gamma)$, as above, with a density function $p(t, x, y)$ that for fixed y satisfy the backward equation (3.12). A rigorous definition about the Diffusion process can be found on page 12 in Bogachev et al. (2015) and also in the same page we have Proposition 1.3.1 which says that the transition probability of such process is the solution of the corresponding Fokker-Planck, see for instance (3.13).

Let now take an Itô diffusion X, the solution of equation (3.4) with coefficients which satisfy the Lipschitz condition. We also know that an Itô diffusions satisfy the strong Markov property, admits an infinitesimal generator and characteristic operator, see for instance Chapter 7 in Øksendal (2010). Let take a function $f \in C_c^2(\mathbb{R})$ and apply Itô's formula, see for instance Theorem 6.2 in Mao (2007)

$$f(X_t) - f(x) = \int_0^t f'(X_s)\mu(X_s) + \tfrac{1}{2}f''(X_s)\sigma^2(X_s)ds + \int_0^t f'(X_s)\sigma(X_s)dW_s$$

Let denote by L, the following differential form

$$L(\cdot) = \mu(x)\frac{d}{dx}(\cdot) + \frac{1}{2}\sigma^2(x)\frac{d^2}{dx^2}(\cdot)$$

Let consider again the semigroup T_t, as in the preceding section. In this case, the limit in the definition of the infinitesimal generator understood point-wise, with respect to x, and the domain $D(A)$ is the set of f where the limit exist for every $x \in \mathbb{R}$. Let consider the following abbreviation $T_tf(x) = \mathbb{E}_xf(X_t)$[4] and proceed with the calculation of the infinitesimal generator

$$Af(x) = \lim_{t\downarrow 0} \tfrac{1}{t}\,\mathbb{E}_x(f(X_t) - f(X_0)) = \lim_{t\downarrow 0} \tfrac{1}{t}\,\mathbb{E}_x(\int_0^t Lf(X_s)ds) = Lf(x)$$

In the above calculation, we use that the expectation of the above stochastic integral is zero, the continuity and the boundedness of $Lf(x)$, which comes from $f \in C_c^2(\mathbb{R})$. Using the fundamental theorem of calculus combined with the dominated convergence theorem, we obtain the above limit.

Notice that A is an extension of the differential operator L, also as we mention in the previous section $f \in D(A)$, $(C_c^2(\mathbb{R}) \subset D(A))$ then $u(t, x) := \mathbb{E}_xf(X_t)$ is the unique

[4]The abbreviation is a shortening for the conditional expectation, $\mathbb{E}_xf(X_t) := \mathbb{E}\,(f(X_t)|X_0 = x)$.

solution of (3.14) within the space of bounded functions, therefore if there is a classical solution for the problem

$$\frac{\partial u(t,x)}{\partial t} = Lu(t,x), \qquad x \in \mathbb{R}, t > 0$$
$$u(0,x) = f(x), \qquad x \in \mathbb{R}$$

it holds that $u(t,x) := \mathbb{E}_x f(X_t)$ is its solution.

Similarly, let consider the following Cauchy problem

$$\frac{\partial u(t,x)}{\partial t} = Lu(t,x) + c(x)u(t,x), \qquad x \in \mathbb{R}, t > 0$$
$$u(0,x) = f(x) \qquad x \in \mathbb{R}$$

where $c(x) \leq 0$ bounded uniformly continuous function, and L, as before, with the extra property that its coefficients are bounded.

If we define the following map $\tilde{T}_t f(x) := \mathbb{E}_x f(X_t) e^{\int_0^t c(X_s)ds}$ acting over B. It is easy to see that $\tilde{T}_t f$ is a semigroup. Let us denote by $\tilde{A}$ the infinitesimal generator of $\tilde{T}_t f$. See that Lemma 3.4.1 implies that $\tilde{A}f(x) = Af(x) + c(x)$ for every $f \in C_c^2(\mathbb{R})$. This implies that the solution of the above Cauchy problem can be written as follows

$$u(t,x) := \tilde{T}_t f(x) = \mathbb{E}_x f(X_t) e^{\int_0^t c(X_s)ds}$$

One can also prove that the solution of the non-homogeneous parabolic equation:

$$\frac{\partial v(t,x)}{\partial t} = Lv(t,x) + c(x)v(t,x) + f(x), \qquad x \in \mathbb{R}, t > 0$$
$$v(0,x) = 0$$

it is given in the following form,

$$v(x,t) := \mathbb{E}_x \int_0^t f(X_s) e^{\int_0^s c(X_u)du} ds$$

Before, we derive the solution of elliptic problem, let us go back to the definition of a diffusion process. According to Feller (1954a), a process is called of diffusion type, if it is a Markov process which satisfy Kolmogorov backward equation (3.12) and also, if we define the Laplace Transformation over the semigroup $T_t f$

$$F_\lambda(x) = \int_0^\infty e^{-\lambda t} T_t f(x)\, dt, \qquad \lambda > 0$$

is a solution of the differential equation,

$$\lambda F_\lambda(x) - LF_\lambda(x) = f(x)$$

Taking this into account, we can now move on the elliptic problem. Also, in what follows, we present examples which are defined on an open bounded interval D with boundary ∂D.

Example(Dirichlet problem):

$$Lu(x) - c(x)u(x) = f(x), \qquad x \in D$$
$$u(x) = \psi(x), \qquad x \in \partial D$$

where $c(x)$, $f(x)$ and $\psi(x)$ are bounded continuous functions and $c(x)$, $\psi(x)$ are non-negative. Let us also consider operator L with coefficients which satisfy Lipschitz condition and σ has a positive lower bound for $x \in \overline{D}$, closure of D. Next, let us define the first exit from the boundary with $\tau = \inf\{t \geq 0 : X_t \notin D\}$, if D is open then we have that τ is a stopping time, see for instance Example 7.2.2. in Øksendal (2010).

Thus it is obvious that, we should start by applying the Itô formula to $u(X_t)e^{Y_t}$, where $Y_t = -\int_0^t c(X_s)ds$ and u is a solution of Dirichlet problem. See also that the process Y_t is monotone, so using a similar version to Theorem 7.14 in Mörters and Peres (2010), we get that

$$u(X_t)e^{Y_t} - u(x) = \int_0^t e^{Y_s}u'(X_s)\sigma(X_s)dW_s + \int_0^t e^{Y_s}\left(Lu(X_s) - u(X_s)c(X_s)\right)ds.$$

Next, we have that for $t < \tau$, it holds that $Lu(X_s) - u(X_s)c(X_s) = f(X_s)$. Also, we have that under the above conditions that $\mathbb{E}_x\tau < \infty$ and according to the definition 5.15 in Mao (2007), we have that

$$\mathbb{E}_x \int_0^{t\wedge\tau} e^{Y_s}u'(X_s)\sigma(X_s)dW_s = 0$$

Therefore, we have that u satisfies the following equation

$$u(x) = \mathbb{E}_x\psi(X_\tau)e^{-\int_0^\tau c(X_s)ds} - \mathbb{E}_x\int_0^\tau f(X_s)e^{-\int_0^s c(X_s)ds}ds.$$

Let us consider a simpler elliptic equation with Neumann boundary conditions. Suppose that $\lambda > 0$ and ψ and L as in the last example,

$$Lu - \lambda u = 0, \qquad x \in D$$
$$u'(x) = \psi(x), \qquad x \in \partial D$$

has a unique solution. Next, let us take the reflected SDE, see for instance (3.11),

$$d\xi_t = \mu(\xi_t)dt + \sigma(\xi_t)dW_t + dl_s, \qquad \xi_0 = x$$

Considering the function $e^{-\lambda t}u(x)$ and applying Itô's formula, we get that

$$\begin{aligned}\mathbb{E}_x u(\xi_t)e^{-\lambda t} - u(x) &= \mathbb{E}_x\int_0^t e^{-\lambda s}(L-\lambda I)u(\xi_s)ds + \int_0^t e^{-\lambda s}u'(\xi_s)dl_s \\ &= \mathbb{E}_x\int_0^t e^{-\lambda s}\psi(\xi_s)dl_s\end{aligned} \tag{3.15}$$

For the last equality, we use part (iii) in definition 3.2.2, which provide us that ξ_s does not sticky on the boundary for positive time, therefore, we can use that $(L - \lambda) u = 0$. For the following we use the boundedness of u, similarly with the last example u assumed to be the solution of the elliptic problem on a bounded interval. According to the last equality, we have that $u(x)$ satisfies the above equation for each t, thus, by letting t goes to infinite, and applying the Dominated convergence theorem on the right-hand side and the monotone convergence theorem on the left hand-side we have that

$$u(x) = -\mathbb{E}_x \int_0^\infty e^{-\lambda s} \psi(\xi_s) dl_s$$

Let us now take an open bounded interval D, the boundary consists of two components ∂D_1 and ∂D_2, also ψ and λ are defined as above,

$$\begin{aligned} &Lu - \lambda u = 0, && x \in D \\ &u(x) = 0, && x \in \partial D_1 \\ &u'(x) = \psi(x), && x \in \partial D_2 \end{aligned}$$

Let u has a unique solution u. Let again consider the first exit from the boundary ∂D_1 and denote it by τ. Similarly, we get that that equation (3.15) holds also for this case, so we have that

$$\mathbb{E}_x u\left(\xi(t \wedge \tau)\right) e^{-\lambda(t \wedge \tau)} - u(x) = \mathbb{E}_x \int_0^{t \wedge \tau} e^{-\lambda s} \psi(\xi_s) dl_s$$

Similar to the previous example, we get that

$$u(x) = \mathbb{E}_x u\left(\xi(\tau)\right) e^{-\lambda \tau} - \mathbb{E}_x \int_0^\tau \psi(\xi_s) dl_s$$

The last example is for mixed boundary value problem, for λ and ψ as above, we consider the problem,

$$\begin{aligned} Lu &= 0, && x \in D \\ u' - \lambda u &= g, && x \in \partial D \end{aligned}$$

Let u be the unique solution and define

$$Y_t = u(\xi_t) - \int_0^t u'(\xi_s) dl_s$$

Y_t is martingale[5], next apply Itô's formula for the product $Y_t e^{-\lambda l_t}$,

$$d\left(Y_t e^{-\lambda l_t}\right) = e^{-\lambda l_t}\left(u'(\xi_t)\sigma(\xi_t) dW_t - \lambda Y_t dl_t\right)$$

[5]Take into account that u is the solution of the elliptic problem and by applying Itô's formula on Y_t, we get that $dY_t = u'(\xi_t)\sigma(\xi_t) dW_t$

equivalently,

$$d\left(u(\xi_t)e^{-\lambda l_t}\right) - d\left(e^{-\lambda l_t}\int_0^t u'(\xi_s)dl_s\right)$$
$$= e^{-\lambda l_t}\left(u'(\xi_t)\sigma(\xi_t)dW_t - \lambda u(\xi_t)dl_t + \lambda\int_0^t u'(\xi_s)dl_s\, dl_t\right)$$

equivalently[6],

$$d\left(u(\xi_t)e^{-\lambda l_t}\right) = \left(u'(\xi_t) - \lambda u\right)(\xi_t)\, e^{-\lambda l_t}dl_t + e^{-\lambda l_t}u'(\xi_t)\sigma(\xi_t)dW_t$$

Similar as before, we have

$$u(\xi_t)e^{-\lambda l_t} - u(\xi_0)e^{-\lambda l_0} = \int_0^t \psi(\xi_s)e^{-\lambda l_s}dl_s + \int_0^t e^{-\lambda l_s}u'(\xi_t)\sigma(\xi_t)dW_t \tag{3.16}$$

Under the assumption that σ is uniformly positive, we can show that: $l_t \to \infty$ as $t \to \infty$ and so to obtain that

$$\lim_{t\to\infty} \mathbb{E}_x e^{-\lambda l_t} = 0 \tag{3.17}$$

In order to prove it, one can for instance use equation (21) in Anderson and Orey (1976), in that paper, the authors consider an open set U_0 inside the interval D and defines two sequences of stopping times: $S_0 = 0$, $S_1 = \inf\{t : X \in U_0\}$, $T_1 = \inf\{t > S_1 : X \in \partial D\}$, $\ldots S_k = \inf\{t > T_{k-1} : X \in U_0\}$, $T_k := \inf\{t > S_k : X \in \partial D\}$, then shows that there exists positive c such that for every $k \geq 1$ the following hold

$$\mathbb{P}\left(l_{S_{k+1}} - l_{S_k} > c|\mathcal{F}_{S_k}\right) > \frac{1}{4}$$

Using the last inequality, the authors prove that for every $M > 0$ and every $n \geq 8M/c$, one can conclude that:

$$\mathbb{P}\left(\sum_{k=1}^{n}(l_{S_{k+1}} - l_{S_{k+1}}) < M\right) \to 0, \qquad \text{as } n \to \infty$$

from that point, it is easy to obtain our claim.

Next, we apply the expectation on the formula (3.16) and take the limits for t to ∞, using that u is bounded and the limit (3.17), we get that:

$$u(x) = -\mathbb{E}_x\int_0^\infty \psi(\xi_t)e^{-\lambda l_t}dl_t$$

3.4 Proof for Theorem 3.2.4 and Lemma 3.4.1

Lemma 3.4.1. *Suppose that X is a solution of equation (3.4) with uniformly bounded coefficients μ and σ which satisfy Lipschitz and linear growth properties. Consider a bounded*

[6] We have that: $d\left(e^{-\lambda l_t}\int_0^t u'(\xi_s)dl_s\right) = -\lambda e^{-\lambda l_t}\int_0^t u'(\xi_s)dl_s dl_t + e^{-\lambda l_t}u'(\xi_t)(\xi_t)dl_t$

and uniformly continuous function $c(x) \leq 0$ and define $\tilde{T}_t f(x) := \mathbb{E}_x f(X_t) e^{\int_0^t c(X_s)ds}$ over the set of bounded measurable functions with its infinitesimal generator denoted by $\tilde{A}$. Let us also denote by A the infinitesimal generator of $T_t f(x) := \mathbb{E}_x f(X_t)$. Then, we have that for every $f \in C_c^2(\mathbb{R})$, it holds that

$$\tilde{A} f(x) = A f(x) + c(x) f(x)$$

Proof.

$$\begin{aligned}
\tilde{A} f(x) &= \lim_{t\downarrow 0} \frac{1}{t} \mathbb{E}_x \left(f(X_t) e^{\int_0^t c(X_s)ds} - f(x) \right) \\
&= \lim_{t\downarrow 0} \frac{1}{t} \mathbb{E}_x \int_0^t d\left(f(X_q) e^{\int_0^q c(X_s)ds} \right) \\
&= \lim_{t\downarrow 0} \frac{1}{t} \mathbb{E}_x \int_0^t e^{\int_0^q c(X_s)ds} \left(f(X_q)c(X_q) + f'(X_q)\mu(X_q) + \frac{1}{2} f"(X_q)\sigma^2(X_q) \right) dq \\
&= \lim_{t\downarrow 0} \frac{1}{t} \mathbb{E}_x \int_0^t e^{\int_0^q c(X_s)ds} \left(f(X_q)c(X_q) + Lf(X_q) \right) dq
\end{aligned}$$

Note also that, from $c(x) \leq 0$, we have that $e^{\int_0^q c(X_s)ds} \leq 1$, also since $f \in C_c^2(\mathbb{R})$, we get the continuity and the boundedness of $f(X_q), Lf(X_q)$ and also by definition, we have that $c(x)$ is also bounded.

Using the dominated convergence theorem, we have that

$$\begin{aligned}
\tilde{A} f(x) &= \mathbb{E}_x \lim_{t\downarrow 0} \frac{1}{t} \int_0^t e^{\int_0^q c(X_s)ds} \left(f(X_q)c(X_q) + Lf(X_q) \right) dq \\
&= f(x)c(x) + Lf(x)
\end{aligned}$$

the last equality is a consequence of fundamental theorem of calculus and the continuity $f(X_q)c(X_q) + Lf(X_q)$ with respect to q. □

3.4.1 Uniqueness for reflecting barrier

In this proof, we consider an increasing sequence of $C^2(\mathbb{R})$ functions denoted by $\phi_k(x)$, such that its limit converges in the absolute value of x, i.e. $|\cdot|$. Using that sequence, we can show that for any two (ξ_1, l_1) and (ξ_2, l_2) satisfying equation (3.11), it holds that $\mathbb{P}\left(\xi_1(t) = \xi_2(t), \forall t \geq 0\right) = 1$.

First, let build the sequence of those ϕ_k:

We have that for every $T > 0$, it holds that $\int_0^T 1/u \, du = \infty$. Hence, we can choose a positive decreasing sequence of a_k that goes to zero such that: $\int_{a_k}^{a_{k-1}} 1/u \, du = k$. Let take a sequence of $\phi_k \in C^2\left(\mathbb{R}^+\right)$ with the following properties:

$$\phi_k(0) = 0, \qquad \phi_k(u)' := \begin{cases} 0, & u \in [0, a_k] \\ [0,1], & u \in (a_k, a_{k-1}), \\ 1, & u \in [a_{k-1}, \infty) \end{cases} \qquad \phi_k(u)'' := \begin{cases} 0, & u \in [0, a_k] \\ [0, \frac{2}{ku}], & u \in (a_k, a_{k-1}) \\ 0, & u \in [a_{k-1}, \infty) \end{cases}$$

Once, we have ϕ_k for $x \geq 0$, they can be extended on $\mathbb{R}$ in the following way: $\phi_k(u) = \phi_k(|u|)$. Therefore, it is easy to see that: $\phi_k'(u) = -\phi_k'(-u)$ and $\phi_k(u) \nearrow |u|$.

Since, ϕ_k are twice differentiable, we can apply Itô's formula, see for instance Theorem 6.2 at page 32 Mao (2007),

$$\phi_k(\xi_1(t) - \xi_2(t)) = \int_0^t \phi_k'(\xi_1 - \xi_2)(\mu(\xi_1) - \mu(\xi_2))ds + \frac{1}{2}\int_0^t \phi_k''(\xi_1 - \xi_2)(\sqrt{\sigma(\xi_1)} - \sqrt{\sigma(\xi_2)}\,)^2\,ds$$
$$+ \int_0^t \phi_k'(\xi_1 - \xi_2)(\sqrt{\sigma(\xi_1)} - \sqrt{\sigma(\xi_2)}\,)dW_s + \int_0^t \phi_k'(\xi_1 - \xi_2)d(l_1 - l_2)$$

Let us denote the above integrals by $I_1, \dots I_4$, respectively. Using the properties of ϕ_k and also Lipschitz conditions of μ and σ, we get

$$I_1 \leq K_\mu \int_0^t |\xi_1 - \xi_2| ds$$
$$I_2 \leq \frac{1}{k}\int_0^t 1(|\xi_1 - \xi_2| \in (a_k, a_{k-1}))\frac{|\sigma(\xi_1) - \sigma(\xi_2)|}{|\xi_1 - \xi_2|}ds \leq \frac{1}{k}K_\sigma\, t$$

According to the definition of l_i, see part (iii) 3.2.1, we have the last integral

$$I_4 = \int_0^t 1(\xi_1 = 0)\phi_k'(\xi_1 - \xi_2)dl_1(s) - \int_0^t 1(\xi_2 = 0)\phi_k'(\xi_1 - \xi_2)dl_2(s)$$
$$= -\left(\int_0^t 1(\xi_1 = 0)\phi_k'(\xi_2)dl_1(s) + \int_0^t 1(\xi_2 = 0)\phi_k'(\xi_1)dl_2(s)\right) \leq 0$$

The last equality holds because ϕ_k' is an odd function. Also, the last part is non-positive because ξ_i are non-negative, thus $\phi_k'(\xi_i)$ are non-negative, furthermore l_i are increasing functions, thus the integrals with respect to $dl_i(s)$ over a non-negative function are also non-negative.

Let fix n and denote the stopping time:

$$\tau_n = \inf\{t \geq 0 : |\xi_1| \wedge |\xi_2| \geq n\}$$

So, we have

$$\mathbb{E}\,\phi_k(\xi_1(t \wedge \tau_n) - \xi_2(t \wedge \tau_n)) \lesssim \mathbb{E}\int_0^{t\wedge\tau_n} |\xi_1 - \xi_2|(s)\,ds + \frac{1}{k}\,\mathbb{E}\,t \wedge \tau_n$$
$$+ \mathbb{E}\int_0^{t\wedge\tau_n} \phi_k'(\xi_1 - \xi_2)(\sqrt{\sigma(\xi_1)} - \sqrt{\sigma(\xi_2)}\,)dW_s$$

We have that: $\int_0^t \phi_k'(\xi_1 - \xi_2)(\sqrt{\sigma(\xi_1)} - \sqrt{\sigma(\xi_2)}\,)1(s < \tau_n)dW_s$ is martingale, so the last term is zero. In this point, we need to mention that the proportionality of the above inequality does not depend on the choice of k. Therefore, using the monotonicity of ϕ_k

$$\mathbb{E}|\xi_1 - \xi_2|(t \wedge \tau_n) = \lim_k \mathbb{E}\phi_k\left(\xi_1(t \wedge \tau_n) - \xi_2(t \wedge \tau_n)\right) \lesssim \limsup_k \left(\mathbb{E}\int_0^{t\wedge\tau_n} |\xi_1 - \xi_2|(s)\,ds + \frac{t \wedge \tau_n}{k}\right)$$
$$= \mathbb{E}\int_0^{t\wedge\tau_n} |\xi_1 - \xi_2|(s)\,ds \leq \int_0^t \mathbb{E}|\xi_1 - \xi_2|(s \wedge \tau_n)\,ds$$

Using Grönwall's inequality, one can see that:

$$\mathbb{E}|\xi_1 - \xi_2|(t \wedge \tau_n) = 0, \quad \forall t \geq 0$$

Therefore, we have that: $\forall t \geq 0$ and n, it holds $\mathbb{P}\left(\, \xi_1(t \wedge \tau_n) = \xi_2(t \wedge \tau_n) \,\right) = 1$, the monotonicity of τ_n implies that: $\forall t \geq 0$, it holds $\mathbb{P}\left(\, \xi_1(t) = \xi_2(t) \,\right) = 1$. Last, the continuity of ξ_i provides us the following:

$$\begin{aligned}\mathbb{P}\left(\xi_1(t) = \xi_2(t), \forall t \geq 0\right) &= \mathbb{P}\left(\xi_1(t) = \xi_2(t), \forall t \in \mathbb{Q}_+\right) \\ &= \lim_{m \to \infty} \mathbb{P}\left(\xi_1(t) = \xi_2(t), \forall t \in \mathcal{P}_m\right) = 1\end{aligned}$$

where $\mathcal{P}_m$ is an increasing sequence of partition of $\mathbb{Q}_+$ and m is the cardinality of that set. The first equality holds because $\mathbb{Q}_+$ is dense in $\mathbb{R}_+$ and a.e. path of ξ_i is continuous, where for the second equality use $\mathcal{P}_m$ is an increasing sequence and also holds that $\mathbb{P}\left(\xi_1(t) = \xi_2(t), \forall t \in \mathcal{P}_m\right) = 1$ for every m.

Last, the uniqueness of l_i follows from equation (3.11)

$$l_1(t) = \xi_1(t) - \xi_0 - \int_0^t a(\xi_1(s))ds - \int_0^t b(\xi_1(s))dW_s = l_2(t), \ \forall t \geq 0 \, \text{a.s.}$$

Chapter 4

Dimension reduction for exponential priors

4.1 Introduction

In this chapter, we consider probabilities μ and π with the assumption that the Radon-Nikodym derivative of μ with respect to π, similar to Section 2.6.1, is given, up to a multiplication constant. Here μ and π are defined on a finite but high-dimensional space $\mathbb{R}^d$. We consider a methodology for the identification of the optimal subspace for sampling in the case of exponential priors. That methodology was first proposed in Zahm et al. (2021), and recovers a subspace that contains the most weight of the prior distribution based on Kullback-Leibler divergence. The methodology is based on the logarithmic Sobolev inequality. The logarithmic Sobolev inequality can be obtained for Gaussian measures, and then can be extend also in the of sub-Gaussians. Therefore, the methodology depends on the prior distribution.

Observe that 1-exponential probabilities measures have heavier tails, i.e. the tails tend to 0 slower than the tails of Gaussian and also that logarithmic Sobolev inequality does not apply to 1-exponential probabilities measures. Therefore, the extension of that methodology is based on a modified logarithmic Sobolev inequality, see in LEDOUX (1997). My main contribution for this extension is Theorem 4.4.3, which allows us to extend the methodology proposed in Zahm et al. (2021) in the case of 1-exponential prior measures. Then the identification and recovery of the optimal subspace come from Proposition 4.4.1, where we can also find the error estimates of that method with respect to Kullback Leibler divergence. In addition, using the proof in LEDOUX (1997), we obtain a local version of the modified logarithmic Sobolev inequality. That local version aims to provide some further control of the method based on the local properties of the likelihood.

4.2 The general methodology

Suppose, we have a probability π which is defined on $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$ and a non-negative integrable function f with respect to π, i.e. $f \in L^1(\pi) \cap \{f \geq 0\}$. Also, we have a probability μ which is defined through the following expression:

$$\frac{d\mu}{d\pi}(u) \propto f(u) \tag{4.1}$$

Observe that the above expression follows the same framework as in Section 2.6.1.

We first review the methodology suggested by Zahm et al. (2021) for approximation of μ on an appropriate subspace of dimension $r < d$. The core idea of this method is the approximation of μ using probabilities which are defined as follows,

$$\frac{d\mu_r}{d\pi}(u) \propto g \circ P_r(u) \tag{4.2}$$

where $P_r : \mathbb{R}^d \to \mathbb{R}^d$ is a linear projection with rank r, i.e. $Im(P_r) \cong \mathbb{R}^r$ which satisfies the property $P_r^2 = P_r$, and $g : \mathbb{R}^d \to \mathbb{R}_+$ is a Borel function. At this point, it is worth mentioning that P_r is not necessarily an orthogonal matrix. Let us now observe the following decomposition: for every $x \in \mathbb{R}^d$, we have that

$$x = (I_d - P_r)x + P_r x = x_\perp + x_r$$

where $x_\perp \in ker(P_r)$ and $x_r \in Im(P_r)$, i.e. $\mathbb{R}^d = ker(P_r) \oplus Im(P_r)$. This method aims to identify an r dimensional subspace of $\mathbb{R}^d$, i.e. $Im(P_r)$, which contains 'most' of the weight of the measure μ. Intuitively speaking, the structure of P_r implies that probability μ_r replaces the function f with a function in the form of $g \circ P_r(x) = g(x_r)$, where $x_r \in Im(P_r) \cong \mathbb{R}^r$.

Therefore, we are looking for a probability μ^{g,P_r} which minimises the distance to μ. The suggested distance for this methodology is the Kullback-Leibler divergence, denoted by $D_{KL}(\cdot||\cdot)$ and it is defined as follows

$$D_{KL}(\mu||\pi) = \int \log\left(\frac{d\mu}{d\pi}\right) d\mu. \tag{4.3}$$

Let us now denote by $\mathcal{P}_r$ the collection of candidate projections P_r

$$\mathcal{P}_r = \left\{P_r \in \mathbb{R}^{d\times d} \,\middle|\, Im(P_r) \cong \mathbb{R}^r, P_r^2 = P_r\right\} \tag{4.4}$$

Then the minimisation problem can be written in the following form:

$$\mu_r^* = \underset{\mu_r \in \mathcal{M}_{\mathcal{P}_r}}{\operatorname{argmin}}\, D_{KL}(\mu||\mu_r)$$

where $\mathcal{M}_{\mathcal{P}_r}$ is defined as follows

$$\mathcal{M}_{\mathcal{P}_r} = \left\{ \mu_r \text{ defined on } \mathcal{B}\left(\mathbb{R}^d\right) \left| \begin{array}{c} g : \mathbb{R}^d \to \mathbb{R}_+ \text{ is a Borel function,} \\ P_r \in \mathcal{P}_r,\ \frac{d\mu_r}{d\pi}(u) \propto g \circ P_r \end{array} \right. \right\}$$

The above minimisation problem can also be written in terms of $g \circ P_r$ as follows

$$(g \circ P_r)^* = \underset{\mu_r \in \mathcal{M}_{\mathcal{P}_r}}{\operatorname{argmin}} D_{KL}\left(\mu || \mu_r\right). \tag{4.5}$$

4.3 Optimal g

In this section, we fix the projection P_r and find the solution of problem (4.5) only with respect to g.

The notation $\sigma(P_r)$ is used for the $\sigma-$algebra generated by P_r, or in other words, is defined as the smallest $\sigma-$algebra which includes the collection $\left\{P_r^{-1}A :\ \text{for some } A \in \mathcal{B}\left(\mathbb{R}^d\right)\right\}$. According to Lemma 2.2 in Zahm et al. (2021), which is a special case of Doob-Dynkin's Lemma, for a given projection $P_r \in \mathbb{R}^{d\times d}$, the following two assertions hold:

$$g \circ P_r \text{ is a } \sigma\left(P_r\right)\text{-measurable function} \tag{4.6}$$

and conversely, for a given $\sigma(P_r)$-measurable function h, there exists a measurable function g defined on $\mathbb{R}^d$ such that

$$h = g \circ P_r \tag{4.7}$$

for the general statement of Doob-Dynkin's Lemma, see for instance Lemma 1.13 in Kallenberg and Kallenberg (1997). Essentially, the Lemma allows us to replace $g \circ P_r$ with any $\sigma(P_r)$-measurable function. Hence, our first consideration is on the conditional expectation of f given $\sigma(P_r)$ under the distribution π, denoted by $\mathbb{E}_\pi\left(f|\sigma\left(P_r\right)\right)$, that defines a real valued function on $\mathbb{R}^d$. A property of the conditional expextation is that $\mathbb{E}_\pi\left(f|\sigma\left(P_r\right)\right)$ is the unique $\sigma(P_r)$ measurable function which for every $\sigma(P_r)$-measurable function h, satisfies

$$\int f\, h\, d\pi = \int \mathbb{E}_\pi\left(f|\sigma\left(P_r\right)\right)\, h\, d\pi. \tag{4.8}$$

We can now state the following result by Zahm et al. (2021).

Lemma 4.3.1. *Suppose P_r and consider the following minimisation problem*

$$\underset{g \text{ Borel function}}{\operatorname{argmin}} \quad D_{KL}\left(\mu || \mu^{g\circ P_r}\right)$$

Then, we have that the above problem attains its optimal solution at $\tilde{\mu}_r$ given by

$$\frac{d\tilde{\mu}_r}{d\pi}(u) \propto \mathbb{E}_\pi\left(f|\sigma\left(P_r\right)\right). \tag{4.9}$$

Proof. Given μ^{g,P_r} we evaluate the following difference,

$$\begin{aligned} D_{KL}\left(\mu||\mu^{g,P_r}\right) - D_{KL}\left(\mu||\tilde{\mu}_r\right) &= \int \log\left(\frac{Z_{g\circ P_r}}{Z_{\mathbb{E}_\pi(f|\sigma(P_r))}} \frac{\mathbb{E}_\pi(f|\sigma(P_r))}{g\circ P_r}\right) \frac{f}{Z_f} d\pi \\ &= \int \log\left(\frac{Z_{g\circ P_r}}{Z_f} \frac{\mathbb{E}_\pi(f|\sigma(P_r))}{g\circ P_r}\right) \frac{\mathbb{E}_\pi(f|\sigma(P_r))}{Z_f} d\pi = D_{KL}\left(\tilde{\mu}_r||\mu^{g,P_r}\right) \end{aligned}$$

where $Z_{g\circ P_r}$, $Z_{\mathbb{E}_\pi(f|\sigma(P_r))}$, and Z_f are the normalisation constants. Using (4.8) we have

$$Z_f := \int f d\pi = \int \mathbb{E}_\pi\left(f|\sigma\left(P_r\right)\right) d\pi = Z_{\mathbb{E}_\pi(f|\sigma(P_r))}.$$

The second equality follows from (4.8), observe that the logarithm is a $\sigma\left(P_r\right)$-measurable function. According to (2.12), we have that the Kullback–Leibler divergence is non-negative, so using it in the above difference, we have that

$$D_{KL}\left(\mu||\tilde{\mu}_r\right) \leq D_{KL}\left(\mu||\mu^{g,P_r}\right)$$

giving the result. □

4.3.1 Properties of conditional expectation on P_r

We start with showing that the sets in the collection $\left\{P_r^{-1}A : \text{ for some } A \in \mathcal{B}\left(\mathbb{R}^d\right)\right\}$ have the following form

$$B + ker(P_r), \qquad \text{where } B \subseteq Im(P_r). \tag{4.10}$$

Consider $x_r \in Im(P_r)$, $x_\perp \in ker(P_r) \setminus \{0\}$. We have

$$P_r^{-1}\{0\} = ker(P_r), \qquad P_r^{-1}\{x_r\} = ker(P_r) + \{x_r\}, \qquad P_r^{-1}\{x_\perp + x_r\} = \emptyset.$$

Notice that the above formula for $\{x_r\}$ comes from the property $P_r^2 = P_r$. Hence, every set from the above collection is written in the form of (4.10). In addition, we have that $\sigma(P_r)$ is the smallest $\sigma-$algebra generated by sets of the form (4.10).

Therefore, it is natural to construct a map U which separates $\mathbb{R}^d$ in the following two subspace $Im(P_r)$ and $ker(P_r)$. Thus, we consider the invertible matrix $U = (U_r|U_\perp)$: $\mathbb{R}^r \times \mathbb{R}^{d-r} \to \mathbb{R}^d$, such that $U(\mathbb{R}^r \times \{0\}) = Im(P_r)$ and $U(\{0\} \times \mathbb{R}^{d-r}) = ker(P_r)$. Based on that matrix, Proposition 2.4 in Zahm et al. (2021) provides a representation for the conditional expectation of f given $\sigma(P_r)$ under the probability measure π. We recall it here.

Proposition 4.3.1. *For any probability measure π which admits a probability density function ρ, i.e. $\pi(dx) = \rho(x)dx$, and for any r-rank projector P_r, consider the matrix*

$U_{\perp} \in \mathbb{R}^{d\times(d-r)}$ *with columns consisting of vectors that form a basis on ker*(P_r). *Let* $p_{\perp}$ *be the conditional probability density on* $\mathbb{R}^{d-r}$, *defined by*

$$p_{\perp}\left(\xi_{\perp}|P_r x\right) = \frac{\rho\left(P_r x + U_{\perp}\xi_{\perp}\right)}{\int_{\mathbb{R}^{d-r}} \rho\left(P_r x + U_{\perp}\xi'_{\perp}\right) \mathrm{d}\xi'_{\perp}} \tag{4.11}$$

for every $\xi_{\perp} \in \mathbb{R}^{d-r}$ *and* $x \in \mathbb{R}^d$, *under the convention that* $p_{\perp}\left(\xi_{\perp}|P_r x\right) = 0$ *whenever the denominator of* (4.11) *is zero. Then, for any Borel function* f, *the conditional expectation* $\mathbb{E}_{\mu}\left(f|\sigma\left(P_r\right)\right)$ *is given us as follows*

$$\mathbb{E}_{\pi}\left(f|\sigma\left(P_r\right)\right)(x) := \int_{\mathbb{R}^{d-r}} f\left(P_r x + U_{\perp}\xi_{\perp}\right) p_{\perp}\left(\xi_{\perp}|P_r x\right) \mathrm{d}\xi_{\perp}.$$

Remark 4.3.1. *i) Looking at the proof of the above Proposition (Zahm et al. (2021)), one can define the following integration rule that applies to every Borel function* f*:*

$$\begin{aligned} f(x)\,\pi(dx) &= f(x)\,\rho(x)dx = f\circ U(\xi)\,|U|\rho(U\xi)d\xi \\ &= f\circ U(\xi)\,\frac{\rho(U_r\xi_r + U_{\perp}\xi_{\perp})}{\int \rho(U_r\xi_r + U_{\perp}\xi'_{\perp})d\xi'_{\perp}}d\xi_{\perp}\,\left(|U|\int \rho(U_r\xi_r + U_{\perp}\xi''_{\perp})d\xi''_{\perp}\right)d\xi_r \\ &= f\circ U(\xi)\,p_{\perp}(\xi_{\perp}|U_r\xi_r)d\xi_{\perp}\;p_r(\xi_r)d\xi_r = f\circ U(\xi)\,\pi_{\perp}(d\xi_{\perp}|\xi_r)\,\pi_r(d\xi_r) \end{aligned} \tag{4.12}$$

where ξ_r, $\xi_{\perp}$ *take values in* $\mathbb{R}^r$ *and* $\mathbb{R}^{d-r}$, *respectively,* $\xi = \begin{pmatrix} \xi_r \\ \xi_{\perp} \end{pmatrix}$ *and the probabilities* $\pi_{\perp}(\cdot|\xi_r)$, $\pi_r(\cdot)$ *are defined through the probability density functions* $p_{\perp}(\cdot|U_r\xi_r)$, $p_r(\cdot)$ *on the following Borel* σ*-algebras* $\mathcal{B}\left(\mathbb{R}^{d-r}\right)$ *and* $\mathcal{B}\left(\mathbb{R}^r\right)$, *respectively.*

Since P_r *is a projection hence* $P_r^2 = P_r$, *and by definition of* U_r *and* $U_{\perp}$, *which says that every column vector of* U_r *and* $U_{\perp}$ *form a basis for* $Im(P_r)$ *and* $Ker(P_r)$ *respectively, we get that* $P_r U = P_r U_r = U_r$. *Therefore, the above mentioned representation of* $\mathbb{E}_{\pi}\left(f|\sigma\left(P_r\right)\right)$ *can be written as follows*

$$\mathbb{E}_{\pi}\left(f|\sigma\left(P_r\right)\right)\left(U_r\xi_r + U_{\perp}\xi'_{\perp}\right) := \int_{\mathbb{R}^{d-r}} f\left(U_r\xi_r + U_{\perp}\xi_{\perp}\right) p_{\perp}\left(\xi_{\perp}|P_r x\right) \mathrm{d}\xi_{\perp} \tag{4.13}$$

ii) Another interesting property of the conditional expectation given a σ*-algebra generated by a projection is that: for any two projections* Q_r, P_r *with ker*$(Q_r) =$ *ker*(P_r), *the following holds*

$$\mathbb{E}_{\pi}\left(f|\sigma\left(P_r\right)\right) = \mathbb{E}_{\pi}\left(f|\sigma\left(Q_r\right)\right). \tag{4.14}$$

To see that let us consider Q_r, P_r *as above, then we have that* $Q_r \circ P_r = Q_r$ *and* $P_r \circ Q_r = P_r$, *see for instance the proof of Proposition 2.2 in Zahm et al. (2019). Let now use the last observation in combination with* (4.6) *and* (4.7), *we are able to get* *that every $\sigma(P_r)$-measurable function is $\sigma(Q_r)$-measurable, and conversely. Hence, for any h $\sigma(P_r)$-measurable, or $\sigma(Q_r)$-measurable, we have*

$$\int f\, h\, d\pi = \int \mathbb{E}_\pi \left(f|\sigma\left(P_r\right)\right)\, h\, d\pi = \int \mathbb{E}_\pi \left(f|\sigma\left(Q_r\right)\right)\, h\, d\pi$$

the uniqueness of such a function implies that (4.14) holds.

4.4 Constructing P_r

In this section, we complete the theoretical part of the method, by minimising an upper bound on the distance between μ and $\tilde{\mu}_r$ given in (4.2) and (4.9) respectively. More precisely, we first obtain an upper bound on the the Kullback-Leibler divergence of μ and $\tilde{\mu}_r$ in terms of the projection P_r. We then obtain a P_r which minimises this upper bound.

By Lemma 4.3.1 the optimal $g \circ P_r$ is attained at $\mathbb{E}_\pi \left(f|\sigma\left(P_r\right)\right)$ and we hence have

$$D_{KL}\left(\mu||\tilde{\mu}_r\right) = \int \log\left(\frac{f}{\mathbb{E}_\pi \left(f|\sigma\left(P_r\right)\right)}\right) \frac{f}{Z_f}\, d\pi. \tag{4.15}$$

According to (4.13), for the calculation of denominator of the above logarithm, we have that as long as P_r is a projection and π admits a Lebesgue density, we can define the conditional probability density $p_\perp$ as in equation (4.11). More precisely, using the integration rule as it is stated in (4.12), we can define the probability measures $\pi_\perp(d\xi_\perp|\xi_r)$ and $\pi_r(d\xi_r)$ which, for every Borel function f, satisfy the following equality

$$\int_{\mathbb{R}^d} f(x)\pi(dx) = \int_{\mathbb{R}^r}\int_{\mathbb{R}^{d-r}} f(U_r\xi_r + U_\perp\xi_\perp)\pi_\perp(d\xi_\perp|\xi_r)\pi_r(d\xi_r).$$

Applying the integration rule together with the representation of $\mathbb{E}_\pi \left(f|\sigma\left(P_r\right)\right)$, see (4.13), on the above mentioned Kullback Leibler divergence, we get that

$$\begin{aligned} Z_f D_{KL}\left(\mu||\tilde{\mu}_r\right) &= \int f \log\left(\frac{f}{\mathbb{E}_{\pi_{m,B}}\left(f|\sigma\left(P_r\right)\right)}\right) d\pi_{m,B} \\ &= \int\int f\circ U(\xi)\log\left(\frac{f\circ U(\xi)}{\int f\circ U(\tilde{\xi})\pi_\perp(d\tilde{\xi}_\perp|\xi_r)}\right)\pi_\perp(d\xi_\perp|\xi_r)\,\pi_r(d\xi_r) \\ &= \int Ent_{\pi_\perp(\cdot|\xi_r)}\left(f\circ U\right)\pi_r(d\xi_r) \end{aligned} \tag{4.16}$$

where $\tilde{\xi} = \begin{pmatrix} \xi_r \\ \tilde{\xi}_\perp \end{pmatrix}$ and the last equality emerges from the definition of the entropy of $f \circ U$ under the probability $\pi_\perp(d\xi_\perp|\xi_r)$ given below.

Definition 4.4.1. *Suppose a probability π and a non-negative function h with $\mathbb{E}_\pi h \log^+(h) < \infty$. Then we define the* **entropy** *of h under the probability π as follows,*

$$Ent_\pi\left(h\right) := \mathbb{E}_\pi h \log\left(h\right) - \mathbb{E}_\pi h\ \log\left(\mathbb{E}_\pi h\right). \tag{4.17}$$

Therefore, the suggested minimisation of $D_{KL}(\mu||\tilde{\mu}_r)$ is based on some entropy inequality, or more precisely, on the logarithmic Sobolev inequality.

Let us denote the Lebesgue measure with λ. Then, we say that λ satisfies the *logarithmic Sobolev inequality*, if for every smooth function f, the following inequality holds

$$\int f^2(x) \log\left(\frac{f^2(x)}{\int f^2\, d\lambda}\right) \lambda(dx) \leq C \int \|\nabla f(x)\|_2^2 \lambda(dx)$$

The above inequality emerges from Sobolev inequalities, a proof for the Lebesgue measure is given in Gentil (2003), also, we can show that C is independent from the choice of f and the dimension of $\mathbb{R}^d$.

An analogous inequality for Gaussian measures, $\gamma(dx) = \frac{1}{(2\pi)^{n/2}} e^{-\|x\|_2^2} \lambda(dx)$, has been proved in Gross (1975). In addition, paper Zahm et al. (2021) summarises some conditions on the probability π that imply that π satisfies LSI. Those conditions are the following: probability π needs to have convex support $K = \text{supp}(\pi) \subseteq \mathbb{R}^d$ and its density function ρ should satisfy $\rho(x) \propto e^{-(V(x)+\Psi(x))}$, where $V \in C^2(K)$ is a strongly convex function on K and Ψ is bounded in K. More precisely, the above assumptions imply the existence of a positive definite matrix Γ and a constant $\kappa \geq 1$ such that

$$\nabla^2 V(x) - \Gamma, \text{ is a positive semi-definite matrix} \quad \forall x \in \mathbb{R}^d$$

and

$$e^{\sup \Psi - \inf \Psi} \leq \kappa$$

Then, such probabilities satisfy the following inequality for a sufficiently smooth f

$$\int f^2(x) \log\left(\frac{f^2(x)}{\int f^2\, d\pi}\right) \pi(dx) \leq 2\kappa \int \|\nabla f(x)\|_{\Gamma^{-1}}^2 \pi(dx)$$

where Γ^{-1}-norm is defined through the following inner product, $\|x\|_{\Gamma^{-1}} = \left(\Gamma^{-1}x, x\right)$. The above assumptions are the result of the Bakry-Émery Theorem Bakry and Émery (1985); Bobkov and Ledoux (2000); Otto and Villani (2000) and the Holley–Stroock perturbation Lemma Holley and Stroock (1986).

Observe that the last two assumptions for the density function of the prior probability π cover quite a wide class of probabilities, essentially this class contains the Sub-Gaussian distributions, i.e. distributions decaying to zero at least as fast as the Gaussian tails. Examples include uniform and q-Besov measures, whenever $q \geq 2$, see for instance Section 2.3.1.

A very interesting case of Besov priors not satisfying the above assumptions is the case where $q = 1$. These are of interest especially for the field of signal processing because of their edge-preserving and sparsity-promoting properties, see e.g. Leporini and Pesquet (2001); Kolehmainen et al. (2012); Jia et al. (2016); Tan Bui-Thanh (2015); Rantala et al. (2006). The edge-preserving property is useful when the function of interest has spatial inhomogeneity, i.e. it is extremely spiky in some parts of its domain and in other parts is extremely smooth. These priors are said to be sparsity-promoting property since they promotes solutions that under an appropriate expansion, see for instance equation (2.7), can be represented by a small number of coefficients. That is, there is an appropriate basis $\{\phi_i\}_{i=1}^{\infty}$ such that the most likely draws from the prior are described by relatively few non-zero coefficients u_i in basis $\{\phi_i\}_{i=1}^{\infty}$.

As we mention in Section 2.3.1, the construction of 1-Besov prior is based on 1-exponential probabilities. The considerations presented in the previous paragraph motivate our work in this chapter in extending the approximation methodology of Zahm et al. (2021) to the case of 1-exponential probabilities. Let us recall the formula for the density of 1-exponential measures in $\mathbb{R}^d$

$$\pi_{m,B}\left(dx\right) := e^{-\|B^{-1}(x-m)\|_1} \frac{dx}{2^d|det\left(B\right)|} \tag{4.18}$$

where m is the mean value, B is an invertible matrix on $\mathbb{R}^{d\times d}$ and $\|\cdot\|_1$ is the 1-norm defined on $\mathbb{R}^d$, i.e. $\|x\|_1 = \sum_{i=1}^{d}|x_i|$. We note that in the case of 1-Besov priors, as explained in Section 2.3.1, by construction B is a diagonal matrix.

For establishing a logarithmic Sobolev inequality for the 1-exponential probability measures, we follow the approach by LEDOUX (1997). Therein, the author provides a simplified version of the proof of the concentration property for the 1-exponential probability measures which has been previously proved in Talagrand (1991). The concentration property for the 1-exponential probability measure π_{0,I_d} is given as follows: for every Borel set A with $\pi_{0,I_d}(A) \geq \frac{1}{2}$ and for every $r \geq 0$, it holds that

$$\pi_{0,I_d}\left(A + \sqrt{r}B_2 + rB_1\right) \geq 1 - \mathrm{e}^{-\frac{r}{K}}$$

for some $K > 0$, where B_2 is the Euclidean unit ball and B_1 is the l^1 unit ball in $\mathbb{R}^d$, i.e.

$$B_1 = \left\{x \in \mathbb{R}^d : \sum_{i=1}^{d}|x_i| < 1\right\}$$

At this point, it is worth mentioning that a probability measure π which satisfies a logarithmic Sobolev inequalities, also satisfies a concentration property. This can be shown by using the Herbst argument, see for instance section 2.3 LEDOUX (1997).

We state Theorem 4.2 of LEDOUX (1997) here, which provides us with the modified logarithmic Sobolev inequality for the 1-exponential probability measures. For the sake of completeness, we include the proof of that theorem in Section 4.5. In what follows, we call a positive function $h : \mathbb{R}^d \to \mathbb{R}$ *log-Lipschitz*, if there exists a constant L such that every $x, y \in \mathbb{R}^d$

$$|\log(h(x)) - \log(h(y))| \leq L\|x - y\|.$$

We denote the induced operator norms for a matrix A by $\|A\|_1^*$ and $\|A\|_\infty^*$ for l^1 and l^∞ vector norms respectively.

Theorem 4.4.1 (LEDOUX (1997))**.** *Consider the probability $\pi_{m,B}$, as defined in (4.18). Let $h : \mathbb{R}^d \to \mathbb{R}$ be log-Lipschitz and suppose that $K := \|\nabla \log(h)\|_\infty < 1/\|B\|_\infty^*$. Then,*

$$Ent_{\pi_{m,B}}(h) \leq \frac{2}{1 - (\|B\|_\infty^* K)} \int h\|B^T \nabla \log(h)\|_2^2 \, d\pi_{m,B}.$$

Using the ideas of the proof of the above theorem, we also prove another modified version that can be found in the next section. The purpose of the modified version is to address the case where $K := \|\nabla \log(h)\|_\infty$ is very close to $1/\|B\|_\infty^*$. Then, Theorem 4.4.3 is an application of Theorems 4.4.1 and 4.4.2 on a subspace, which essentially provide us a bound for the Kullback-Leibler divergence, through the expression in (4.16).

Due to limitations of the one-exponential probability measures, we need to re-consider the minimisation problem and more specifically the collection of candidate projections. In case of a Gaussian, or a sub-gaussian, measure the collection $\mathcal{P}_{\pi,r}$, as it is stated in (4.4), works perfectly, see for instance Theorem 2.9 in Zahm et al. (2021). In contrast with the Gaussian case, in order to define the minimisation problem for the one-exponential probability measure $\pi_{m,B}$, we need to restrict ourselves to the following sub-collection

$$\overline{\mathcal{P}}_{\pi_{m,B},r} = \left\{ P_r \in \mathbb{R}^{d\times d} \;\middle|\; \begin{array}{l} Im(P_r) \cong \mathbb{R}^r,\ P_r^2 = P_r,\ \exists P \text{ permutation} \\ \text{such that: } ker\,(P_r) = BP\left(\{0\}^r \times \mathbb{R}^{d-r}\right) \end{array} \right\}.$$

Let us highlight two properties that contribute in the definition of the above collection. Firstly, we observe that Gaussian probabilities are invariant under any rotation, where in this case the probabilities are only invariant under permutation, i.e. $\|(x_1, x_2)\|_1 = \|(x_2, x_1)\|_1$. Essentially, the permutation matrix P is used in such way that the conditional probability emerging from the integration rule stated in (4.12), i.e. $\pi_\perp(\cdot|\xi_r)$, is independent from the choice of ξ_r. Secondly, the kernel property that is stated in the above collection can be explained as follows. Let us consider two projections P_r, Q_r, where their kernels coincide, and define probability measures $\tilde{\mu}_r$, $\overline{\mu}_r$ according to formula (4.9) based on the projections P_r, Q_r, respectively. Then, applying the property in (4.14), one can get that $D_{KL}(\mu||\tilde{\mu}_r) = D_{KL}(\mu||\overline{\mu}_r)$.

Therefore, the proposed minimisation problem for the case of 1-exponential probability is summarised as follows,

$$\mu_r^* = \underset{\mu_r \in \mathcal{M}_{\overline{\mathcal{P}}_{\pi_{m,B}},r}}{\operatorname{argmin}} \quad D_{KL}\left(\mu || \mu_r\right)$$

4.4.1 Local versions of LSI for 1-exponential priors

The following two theorems give local versions of the modified LSI. Hence, the solution of the minimisation problem emerging from this section also depends on the properties of a given set C in combination with the function f.

In the following we call a convex set $C \subseteq \mathbb{R}^d$ a cube, if we can define d connected intervals $C_i \subseteq \mathbb{R}$ such that C can be written in the following way

$$C = C_1 \times C_2 \times \cdots \times C_d$$

An example of such a cube on $\mathbb{R}^d$ is the set $C = \left\{x \in \mathbb{R}^d : \|x\|_\infty \leq 1\right\}$. Let us also observe that the above definition is invariant with respect to translations, see for instance that for a given cube $C \subset \mathbb{R}^d$ and a point $x \in \mathbb{R}^d$, we have that $(C - x)$ is cube.

The following theorem provides a bound for the entropy of $1(\cdot \in C)h$ under the 1-exponential probability, the proof of which can be found in Section 4.5 and is an extension of the proof of Theorem 4.2 in LEDOUX (1997).

Theorem 4.4.2 (Local version of Modified LSI)**.** *Consider the probability $\pi_{m,B}$, as defined in (4.18) and a convex set C. Let h be a log-Lipschitz function with positive $c := \inf\left\{h\left(x\right) : x \in \overline{C}\right\}$. Then the following inequality holds*

$$\int_C h \log\left(\frac{h}{\int_C h \; d\pi_{m,B}}\right) \; d\pi_{m,B} \leq \frac{2}{c} \int_C ||B^T \nabla h||_{2,d}^2 \; d\pi_{m,B} + \tilde{M} \; \pi_{m,B}\left(C^c\right)$$

where $M := \sup\left\{h(x) : x \in \overline{C}\right\}$, $\tilde{M} := d \cdot M$. In addition, if $B^{-1}\left(C\right)$ is a cube, we have that $\tilde{M} := M$. In the particular case where $C = \mathbb{R}^d$, the above inequality is given as follows

$$\int h \log\left(\frac{h}{\int h \; d\pi_{m,B}}\right) \; d\pi_{m,B} \leq \frac{2}{c} \int ||B^T \nabla h||_{2,d}^2 \; d\pi_{m,B}$$

Having obtained a local version of the modified Logarithmic Sobolev inequality for the 1-exponential probabilities, we use equation (4.16) and the probability defined in there, $\pi_\perp(d\xi_\perp|\xi_r)$, to get an upper bound for the Kullback Leibler divergence. The following theorem provides us with sufficient assumptions needed in order to be able to apply the modified LSI on the probability $\pi_\perp(d\xi_\perp|\xi_r)$. The result of the following theorem lets us control $D_{KL}\left(\mu||\tilde{\mu}_r\right)$ in a way that allows us to identify the directions in $\mathbb{R}^d$ that contribute the most to the evaluation of the probability μ for a given probability $\pi_{m,B}$. The proof of the following theorem can be found in Section 4.5. In what follows the standard basis of $\mathbb{R}^d$ is represented by $\{e_i\}_{i=1}^d$

Theorem 4.4.3. *Consider the probability $\pi_{m,B}$, as in (4.18) and a projection $P_r \in \overline{\mathcal{P}}_{\pi_{m,B},r}$ and its corresponding permutation P. Let f be a log-Lipschitz function and C a convex set such that $c := \inf\left\{f(x) : x \in \overline{C}\right\}$ is positive.*

i) Assume Λ to be a convex subset of C such that $B^{-1}(\Lambda)$ is a cube. Then

$$\int f \log\left(\frac{f}{\mathbb{E}_{\pi_{m,B}}\left(f|\sigma\left(P_r\right)\right)}\right)\ \pi_{m,B}(dx) \leq \frac{2}{c}\int_C \|U_\perp^T \nabla f(x)\|_{2,d-r}^2\ \pi_{m,B}(dx) + R_{f,C,\pi_{m,B}}$$

where

$$R_{f,C,\pi_{m,B}} := \max\left\{\tilde{M}, M_{C^c}\left(d-r\right)\left(\tfrac{K\|U_\perp\|_1^*}{1-\beta} + \log\left(K\|U_\perp\|_1^* + 1\right)\right), M_{C^c}\log\left(\tfrac{M_{C^c}}{c\tilde{\lambda}}\right)\right\}\ \pi_{m,B}\left(C^c\right)$$

with $U_\perp = BP\left(e_{r+1},\ldots,e_d\right)$, $M_A := \sup\limits_{x\in\overline{A}} f(x)$ for $A \subset \mathbb{R}^d$, $K = \|\nabla\log\left(f\right)\|_\infty$, $\tilde{M} := d\,M_C$, $\beta := \pi_{0,I_{d-r}}(C_{d-r}^T U^{-1}(C\setminus\Lambda - m))$, and $\tilde{\lambda} \geq \min\left\{\pi_{0,1}\left(e_i^T B^{-1}(\Lambda - m)\right)\right\}^{d-r}$. In addition, if $B^{-1}\left(C\right)$ is cube, we choose $\Lambda = C$ and then we have that $\tilde{M} := M$ and the term $\frac{K\|U_\perp\|_1^*}{1-\beta}$ is becoming $K\|U_\perp\|_1^*$. In the particular case where $C = \mathbb{R}^d$, then $R_{f,C,\pi_{m,B}}$ is equal to zero.

ii) If f is log-Lipschitz with $K = \|U_\perp^T\|_\infty^*\,\|\nabla\log\left(f\right)\|_{L^\infty(\pi_{m,B}),\infty} < 1$, then it holds that

$$\int f\log\left(\frac{f}{\mathbb{E}_{\pi_{m,B}}\left(f|\sigma\left(P_r\right)\right)}\right) d\pi_{m,B} \leq \frac{2}{1-K}\int \|U_\perp^T\nabla\log\left(f\right)\|_{2,d-r}^2\, f\ d\pi_{m,B}$$

Observe that the first bound requires f to be bounded, except in the case where $C = \mathbb{R}^d$. On the other hand for the second bound it is not necessary to assume that f is bounded.

4.4.2 Bounding the error of approximated posterior and constructing P_r

The augmented logarithmic Sobolev inequlaities given in Theorem 4.4.3 result in corresponding bounds on $D_{KL}\left(\mu||\tilde{\mu}_r\right)$ given in the following theorem. The proof is given in Section 4.5.

Theorem 4.4.4. *We consider the same assumptions as in Theorem 4.4.3. We also have that for any $P_r \in \overline{\mathcal{P}}_{\pi_{m,B},r}$ there is a permutation matrix P and for that specific P, there exists a bijective function $\sigma_P : \{1,\ldots d\} \to \{1,\ldots d\}$ such that $P = \left(e_{\sigma_P(1)},\ldots,e_{\sigma_P(d)}\right)$. Then the next two bounds on $D_{KL}\left(\mu||\tilde{\mu}_r\right)$ are respectively corresponding to part (i) and (ii) of Theorem 4.4.3.*

i) $D_{KL}\left(\mu||\tilde{\mu}_r\right) \leq tr(Q^T A\, Q)\; + R_{f,C,\pi_{m,B}}$
where $Q = \left(e_{\sigma_P(r+1)}, \ldots, e_{\sigma_P(d)}\right)$ *and* $A = B^T H_C B$ *with* $H_C := \frac{2}{Z_f\, c}\int_C \nabla f (\nabla f)^T d\pi_{m,B}$.

ii) $D_{KL}\left(\mu||\tilde{\mu}_r\right) \leq tr(Q^T A\, Q)$
where $Q = \left(e_{\sigma_P(r+1)}, \ldots, e_{\sigma_P(d)}\right)$ *and* $A = B^T H B$ *with* $H = \frac{2}{1-K}\int \nabla \log\left(f\right)\left(\nabla \log\left(f\right)\right)^T d\mu$.

The result of the following proposition together with the second bound of the above corollary, gives an upper bound for the following minimisation problem,

$$\min_{\mu_r \in \mathcal{M}_{\overline{\mathcal{P}}_{\pi_{m,B}},r}} D_{KL}\left(\mu||\mu_r\right) \leq \min_{\substack{P\in\mathbb{R}^{d\times d}\ \text{permutation}\\ Q=P(e_{r+1},\ldots,e_d)}} tr(Q^T A Q).$$

Indeed it gives a projection P_r which minimises the trace term in the right-hand side.

Proposition 4.4.1. *For the matrix* $A = (a_{ij}) \in \mathbb{R}^{d\times d}$, *we have*

$$\min_{\substack{P\in\mathbb{R}^{d\times d}\ \text{permutation}\\ Q=P(e_{r+1},\ldots,e_d)}} tr(Q^T A Q) = \sum_{i=1}^{d-r} a_{\tau_A(i),\tau_A(i)}$$

where $\tau_A : \{1, \ldots d\} \to \{1, \ldots d\}$ *is a bijective function which satisfies the following property* $a_{\tau_A(1),\tau_A(1)} \leq \cdots \leq a_{\tau_A(d),\tau_A(d)}$.

Furthermore, the projection

$$P_r = I - \sum_{j=1}^{d-r} y_j y_j^T$$

where $y_j = \frac{u_j}{\|u_j\|_2}$ *with* $u_j = v_j - \left(\sum_{i=1}^{j-1} proj_{u_j}(v_i)\right)$, $v_i = Be_{\tau_A(i)}$ *and* $proj_u(v) := \frac{(u,v)}{(u,u)}u$, *is a minimiser of* $tr(Q^T AQ)$.

4.5 Proofs of dimension reduction for exponential priors

Proof of Theorem 4.4.2 The proof is built in the following way: In the first part, we show the result for $\pi_{0,1}$, the one-exponential probability measure on the real-line. Next, in the second part, we extend the inequality to probability measures $\pi_{0,I_{d\times d}}$ defined on a d-dimensional euclidean space as product measure of d copies of i. i. d measures with density $\pi_{0,1}$. As we see below, the entropy under a product measure can be written in terms of the entropy of the individual measures of that product. The last part is a direct application of Integration by Substitution for the extension to general one-exponential probability measure.

For simplicity, we replace the left hand side of the original inequality with the notation of Entropy, see for instance (4.17) and we also use the notation π instead of the probability

$\pi_{(\cdot,\cdot)}$, whenever it is necessary.

Step 1: we consider probability $\pi_{0,1}$ with probability density function $e^{-|x|}/2$ and denote with $\mathcal{I}(\pi_{0,1})$ the following collection of differentiable functions,

$$\mathcal{I}(\pi_{0,1}) := \left\{ h : \mathbb{R} \to \mathbb{R}_+ \;\middle|\; \begin{array}{c} \text{continuous, differentiable almost everywhere} \\ \text{with } \int |h| d\pi < \infty, \; \int |h'| d\pi < \infty, \; \lim_{x \to \pm\infty} h(x) e^{-|x|} = 0 \end{array} \right\}$$

Notice that using the Integration by parts formula, we can easily show that the following expression holds for any $h \in \mathcal{I}(\pi_{0,1})$,

$$\int h \; d\pi_{0,1} = h(0) + \int sgn(x) \, h'(x) \, \pi_{0,1}(dx) \tag{4.19}$$

Since h is log-Lipschitz, we have that h is differentiable almost everywhere, see for instance Theorem 6 on page 281 in Evans (1998). In addition, a necessary condition for $h \in \mathcal{I}(\pi_{0,1})$ is to assume that for sufficient large x, $|\log(h)'|$ to be uniformly strictly bounded below to 1, i.e. $\exists x_0 >> 1$ such that $z := \sup_{|x| \geq x_0} |\log(h(x))'| < 1$. In order to show that we only need to observe that $\max\left\{h(x)e^{-|x|}, \log(h(x))' \, h(x)e^{-|x|}\right\} \lesssim e^{(z-1)|x|}$ with $z - 1 < 0$.

For the local version of this theorem, we consider the interval $C = (a, b)$, since C is not necessarily bounded we have that $a, b \in \mathbb{R} \cup \{\pm\infty\}$. Similar to (4.19), we have that the integral of h over the set C can be written as follows,

$$\int_a^b h \; d\pi_{0,1} = \mathcal{A}_{\pi_{0,1}}(h, a, b) + \int_a^b sgn(x) \, h'(x) \, \pi_{0,1}(dx)$$

where $\mathcal{A}_{\pi_{0,1}}(h, a, b)$ denotes the boundary term that emerges from the integration by part formula. In order to avoid dealing with the term $\mathcal{A}_{\pi_{0,1}}(h, a, b)$ where is necessary, we can use Lemma 4.5.1.

We begin the proof with the estimation of the following entropy,

$$Ent_\pi(1(\cdot \in C) H) = \mathbb{E}_\pi(1(\cdot \in C) H \; \log(H)) - \mathbb{E}_\pi(1(\cdot \in C) H) \; \log(\mathbb{E}_\pi 1(\cdot \in C) H)$$

where $q := \operatorname{argmin}_{x \in \overline{C}} |x|$ and $H(x) := h(x)/h(q)$. In addition, we observe that for every $u \geq 0$, we have that

$$u - 1 \leq u \log(u)$$

By applying the above inequality to the entropy, we get that

$$Ent_\pi(1(\cdot \in C) H) \leq \mathbb{E}_\pi 1(\cdot \in C)(H \log(H) - H + 1) + \pi(C^c)$$

For simplicity, we consider the measure $\pi(\cdot \cap C)$ and use the notation $\mathbb{E}_{\pi(\,\cdot\, \cap C)}(h)$ as an abbreviation for $\mathbb{E}_\pi(1(\cdot \in C) \; h)$. Notice also that $H\log(H) - H + 1 \in \mathcal{I}(\pi_{0,1})$ and

$H(q)\log(H(q)) - H(q) + 1 = 0$, thus we can apply Lemma 4.5.1 on that function,

$$\mathbb{E}_{\pi(\cdot\,\cap C)}\left(H\log\left(H\right) - H + 1\right) \leq \int_C sgn\left(x\right)H'\left(x\right)\log\left(H\left(x\right)\right)\pi\left(dx\right)$$

Using Cauchy-Schwarz inequality on the right hand side of the above inequality, we get that

$$\mathbb{E}_{\pi(\cdot\,\cap C)}\left(H\log\left(H\right) - H + 1\right) \leq ||H'||_{L^2(\pi(\cdot\,\cap C))}||\log\left(H\right)||_{L^2(\pi(\cdot\,\cap C))} \tag{4.20}$$

Similar to the above two inequalities, it is easy to show that $\log\left(H\right)^2 \in \mathcal{I}\left(\pi_{0,1}\right)$ and $\log\left(H(q)\right)^2 = 0$, so we can apply Lemma 4.5.1 on $\log\left(H\right)^2$. Then, repeating the last two steps for the function $\log\left(H\right)^2$, we get that

$$\mathbb{E}_{\pi(\cdot\,\cap C)}\log\left(H\right)^2 \leq 2\;||\log\left(H\right)||_{L^2(\pi(\cdot\,\cap C))}||\log\left(H\right)'||_{L^2(\pi(\cdot\,\cap C))} \tag{4.21}$$

Applying inequality (4.21) to the inequality (4.20), we get that

$$Ent_{\pi(\cdot\,\cap C)}\left(H\right) \leq 2||H'||_{L^2(\pi(\cdot\,\cap C))}||\log\left(H\right)'||_{L^2(\pi(\cdot\,\cap C))} + \pi\left(C^c\right)$$

Since $H(x) := h(x)/h(q)$, we have that

$$Ent_{\pi}\left(1\left(\cdot\in C\right)h\right) \leq 2||h'||_{L^2(\pi(\cdot\,\cap C))}||\log\left(h\right)'||_{L^2(\pi(\cdot\,\cap C))} + h\left(q\right)\pi\left(C^c\right)$$

Using the positivity of c, we obtain that $|\log\left(h(x)\right)'| \leq \frac{|h(x)'|}{c}$ for every $x \in C$. Thus, we have that

$$Ent_{\pi}\left(1\left(\cdot\in C\right)h\right) \leq \frac{2}{c}\int_I \left(h'\right)^2 d\pi + h\left(q\right)\pi\left(C^c\right) \tag{4.22}$$

Notice also that the above inequality holds even in the case where C is unbounded.

Step 2: we consider probability $\pi_{0,I}$ on the product space $\mathbb{R}\times\ldots\mathbb{R}$, this probability can be written as product of probabilities $\pi_{0,1}$ as it follows

$$\pi_{0,I}\left(dx_1\otimes\cdots\otimes dx_d\right) = \pi_1(dx_1)\cdot\ldots\pi_d(dx_d), \qquad \text{where } \pi_i(dx_i) = \pi_{0,1}(dx_i)$$

Since $\pi_{0,I}$ is product measure, the entropy of $\pi_{0,I}$ satisfies the following inequality, see for instance Proposition 2.2 in LEDOUX (1997),

$$Ent_{\pi_{0,I}}\left(h\right) \leq \sum_{i=1}^{d}\mathbb{E}_{\pi_{0,I}}\left(Ent_{\pi_i}\left(h\right)\right) \tag{4.23}$$

where $Ent_{\pi_i}\left(h\right)$ is defined as follows

$$Ent_{\pi_i}\left(h\right)\left(x_1,\ldots,x_{i-1},x_{i+1},\ldots x_d\right) := \int h(\ldots,x_i,\ldots)\log\left(\frac{h(\ldots,x_i,\ldots)}{\int h(\ldots,x_i,\ldots)\pi_{0,1}(dx_i)}\right)\pi_{0,1}(dx_i)$$

The next step is to apply the inequality obtained in the first part of this proof on the above inequality, but first, we need to introduce the following notation that allows us to cut a given convex set C into slices. In particular, those slices are based on the standard basis of $\mathbb{R}^d$, i.e. $e_i = (0, \ldots, 1 \ldots, 0)^T$.

Let us consider the permutation map $g_i : \mathbb{R}^{d-1} \times \mathbb{R} \to \mathbb{R}^d$ defined as follows $g_i(x^i, x_i) = (x^i_1, \ldots, x_i, \ldots x^i_{d-1})$. It is obvious that g_i can be represented by a permutation matrix P. In addition, it is known that every permutation map is an invertible matrix[1], so the inverse map $g_i^{-1}(x) = \left(x^i, x_i\right)$ is well-defined. Using functions g_i, we can define the following two sets

$$C^i = \left\{ x^i \in \mathbb{R}^{d-1} : \exists x_i \in \mathbb{R} \text{ such that } g_i\left(x^i, x_i\right) \in C \right\}$$

and then, for every $x^i \in C^i$ we define

$$C\left(x^i\right) = \left\{x_i \in \mathbb{R} : g_i(x^i, x_i) \in C\right\}$$

For a given set C and using the above notation, we have that the indicator function of C can be written as follows,

$$1\left(x \in C\right) = 1(x^i \in C^i, x_i \in C(x^i))$$

In addition, let us consider the following notation $h_i(x_i)(x^i) := h \circ g(x^i, x_i)$. Then, the entropy $Ent_{\pi_i}\left(h\right)$, which comes from the inequality (4.23), can be rewritten as follows

$$Ent_{\pi_i}\left(h\right)(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots x_d) = Ent_{X_i \sim \pi_{0,1}}\left(h \circ g(x^i, X_i)\right)$$

Using the above notations, we have that the right hand side of (4.23) can be written in the following way, $\mathbb{E}_{\pi_{0,I}}\left(Ent_{\pi_i}\left(h\right)\right) = \mathbb{E}_{g_i(X^i, X_i) \sim \pi_{0,I}}\left(Ent_{\pi_i}(h_i)(X^i)\right)$. Therefore, the entropy of $1\left(\cdot \in C\right) h$ under the probability $\pi_{0,I}$ is given as follows

$$Ent_{\pi_{0,I}}\left(1\left(\cdot \in C\right) h\right) \leq \sum_{i=1}^{d} \mathbb{E}_{g_i(X^i, X_i) \sim \pi_{0,I}} \left\{1\left(X^i \in C^i\right) \, Ent_{\pi_i}(1\left(\cdot \in C\left(X^i\right)\right) h_i)(X^i)\right\}$$

It is easy to see that for a given convex set C and $x^i \in C^i$, the $C\left(x^i\right)$ is an interval. To see that let us consider the set $\tilde{C}(x^i) := \left\{x \in C : \exists x_i \in C(x_i) \text{ such that } x = g_i(x^i, x_i)\right\}$. Since $\tilde{C}(x^i)$ is defined as the intersection of a line and a convex set C, we have that $\tilde{C}(x^i)$ defines a line segment. Moreover, we have that $C(x^i)$ is the projection of the line segment $\tilde{C}(x^i)$ on the axis of e_i. The continuity of the last projection combined with that the $\tilde{C}(x^i)$ defines a line segment, we have that for every $x^i \in C^i$, $C\left(x^i\right)$ is a connected set.

---

[1] See also that g_i can be written as follows $g_i(x^i, x_i) = \begin{pmatrix} e_1^T \\ \vdots \\ e_{i-1}^T \\ e_{i+1}^T \\ \vdots \\ e_d^T \\ e_i^T \end{pmatrix} \begin{pmatrix} x^i \\ x_i \end{pmatrix}$, where e_i is the standard basis

Therefore, we can apply the inequality (4.22) to the entropy appearing in the last inequality. For the application of that inequality, the original q and c will be replaced by $q_{x^i} = \operatorname{argmin}\left\{|x_i| : x_i \in C\left(x^i\right)\right\}$ and $c_{x^i} = \inf\left\{h \circ g_i(x^i, x_i) : x_i \in \overline{C\left(x^i\right)}\right\}$, respectively. See also that for every $x^i \in C^i$, we have that $c_{x^i} \geq c = \inf\left\{h(x) : x \in \overline{C}\right\}$. The application of this gives us the following inequality

$$
\begin{aligned}
&Ent_{\pi_{0,I_d}}\left(1\left(\cdot \in C\right)h\right) \leq \\
&\quad \frac{2}{c}\sum_{i=1}^{d} \mathbb{E}_{g_i(X^i,X_i)\sim\pi_{0,I}}\left\{1\left(X^i \in C^i\right)\mathbb{E}_{Y_i\sim\pi_{0,1}}1\left(Y_i \in C\left(X^i\right)\right)\left(\partial_i h\right)^2\left(g_i(X^i,Y_i)\right)\right\} \\
&\qquad\qquad + \sum_{i=1}^{d} \mathbb{E}_{g_i(X^i,X_i)\sim\pi_{0,I}} 1\left(X^i \in C^i, X_i \notin C\left(X^i\right)\right) h_i \circ g_i\left(X^i, q_{X^i}\right)
\end{aligned}
$$

Notice that q_{x^i} takes values either from the boundary of $C(x^i)$, or it takes zero in case of $\tilde{C}(x^i) \cap \{0\} \neq \emptyset$. Thus, we can bound $h_i \circ g_i\left(X^i, q_{X^i}\right)$, which appears on the right hand side of the above inequality, by $M = \sup\left\{h(x) : x \in \overline{C}\right\}$. We also notice that the first expectation on the right hand side of the above inequality does not involve the random variable X_i. Using that π_{0,I_d} is a product of π_i, we have that

$$
\begin{aligned}
Ent_{\pi_{0,I}}\left(1\left(\cdot \in C\right)h\right) \leq \frac{2}{c}\int_C ||\nabla h||_2^2 \, d\pi_{0,I_d} \\
+ M\sum_{i=1}^{d} \mathbb{E}_{g_i(X^i,X_i)\sim\pi_{0,I}} 1\left(X^i \in C^i, X_i \notin C\left(X^i\right)\right)
\end{aligned}
\tag{4.24}
$$

Next, we want to estimate the sum in the last inequality:

For generic C, the following holds

$$
\mathbb{E}_{(X^i,X_i)\sim\pi_{0,I}} 1\left(X^i \in C^i, X_i \notin C\left(X^i\right)\right) \leq \pi_{0,I}(C^c)
$$

Let us consider the specific case, where C is cube, then it is easy to show that:

$$
\sum_{i=1}^{d} \mathbb{E}_{(X^i,X_i)\sim\pi_{0,I}} 1\left(X^i \in C^i, X_i \notin C\left(X^i\right)\right) \leq \pi_{0,I}(C^c)
$$

See for example that

$$
\begin{aligned}
\mathbb{E}1((X_1,X_2) \notin C_1 \times C_2) &= \mathbb{E}\left(1(X_1 \notin C_1, X_2 \notin C_2) + 1(X_1 \notin C_1, X_2 \in C_2) + 1(X_1 \in C_1, X_2 \notin C_2)\right) \\
&\geq \mathbb{E}\left(1(X_1 \notin C_1, X_2 \in C_2) + 1(X_1 \in C_1, X_2 \notin C_2)\right)
\end{aligned}
$$

It can be easily seen that $\pi_{0,I}$ is invariant under permutations g_i, i.e. $\pi_{0,I} = {}_{g_i\#}\pi_{0,I}$, where ${}_{g\#}\pi_{0,I}(\cdot)$ is the so-called push forward of $\pi_{0,I}$ and it is defined through the following formula ${}_{g\#}\pi_{0,I}(\cdot) = \pi_{0,I} \circ g^{-1}(\cdot)$. Hence, the above observation about $(X^i, X_i) \sim \pi_{0,I}$ can be applied also to the sum in inequality (4.24). Therefore, we get that

$$
Ent_{\pi_{0,I}}\left(1\left(\cdot \in C\right)h\right) \leq \frac{2}{c}\,\mathbb{E}_{\pi_{0,I}} 1\left(\cdot \in C\right)||\nabla h||_2^2 + \tilde{M}\,\pi_{0,I}\left(C^c\right), \tag{4.25}
$$

where $c := \inf\left\{h(x) : x \in \overline{C}\right\}$, $M := \sup\left\{h(x) : x \in \overline{C}\right\}$, $a := \begin{cases} 1, & \text{if } C \text{ is cube} \\ d, & \text{otherwise} \end{cases}$, and $\tilde{M} := a\, M$

Step 3: We now extend the above inequality to probabilities $\pi_{m,B}$ using integration by substitution.

Since B is invertible, we can define the affine map $g\left(x\right) := Bx + m$. Using the Integration by substitution, we have that $\mathbb{E}_{Y\sim\pi_{m,B}} h\left(Y\right) = \mathbb{E}_{X\sim\pi_{0,I}} h \circ g(X)$. Also, the linearity of $g(x) - m$ implies that the inverse image of a convex set is also convex, thus we have that

$$\begin{aligned}
Ent_{\pi_{m,B}}\left(1\left(\cdot \in C\right)h\right) &= Ent_{\pi_{0,I}}\left(1\left(\cdot \in g^{-1}(C)\right)h \circ g\right) \\
&\leq \frac{2}{c}\int_{g^{-1}(C)} ||\nabla\left(h \circ g\right)||_2^2 \, d\pi_{0,I} + \tilde{M}\ \pi_{0,I}\left(g^{-1}(C)\right) \\
&= \frac{2}{c}\int_{g^{-1}(C)} ||B^T\left(\nabla h\right) \circ g||_2^2 \, d\pi_{0,I} + \tilde{M}\ \pi_{m,B}\left(C\right) \\
&= \frac{2}{c}\int_{C} ||B^T\nabla h||_2^2 \, d\pi_{m,B} + \tilde{M}\ \pi_{m,B}\left(C^c\right)
\end{aligned}$$

where $M := \sup\left\{h \circ g(x) : x \in \overline{g^{-1}(C)}\right\} = \sup\left\{h(x) : x \in \overline{C}\right\}$, this holds due to the continuity of g, similar with M, we get that $c = \inf\left\{h(x) : x \in \overline{C}\right\}$, $a := \begin{cases} 1, & \text{if } g^{-1}(C) \text{ is cube} \\ d, & \text{otherwise} \end{cases}$, and $\tilde{M} := a\, M$.

Proof of Theorem 4.4.1 All the following steps are similar to the Proof of Theorem 4.4.2 except some minor changes that needs to be made. The part for $\pi_{0,1}$ is the original proof as it stated in LEDOUX (1997).

Let us start with the probability measure $\pi_{0,1}$. Here, we consider $H(x) = h(x)/h(0)$ and using the inequality $u - 1 \leq u \log\left(u\right)$ for $u \geq 0$, we have that

$$Ent_\pi\left(H\right) \leq \int sgn\left(x\right) H'\left(x\right) \log\left(H\left(x\right)\right) \pi\left(dx\right) \leq \|\sqrt{H}\log\left(H\right)\|_{L^2(\pi)} \|\sqrt{H}\log\left(H\right)'\|_{L^2(\pi)}$$

also for the last inequality, we use Cauchy-Schwarz inequality. Next, an evaluation follows for the term $\|\sqrt{H}\log\left(H\right)\|_{L^2(\pi)}^2$, see that $H\log(H)^2 \in \mathcal{I}\left(\pi_{0,1}\right)$ and $H(0)\log\left(H(0)\right)^2 = 0$, so the integration by part formula stated in (4.19) implies that

$$\begin{aligned}
\mathbb{E}_\pi H\log(H)^2 &= 2\int sgn\left(x\right) H\left(x\right) \left(\log\left(H\left(x\right)\right)\right)' \log\left(H\left(x\right)\right) \pi\left(dx\right) \\
&\qquad + \int sgn\left(x\right) H'\left(x\right) \log\left(H\left(x\right)\right)^2 \pi\left(dx\right) \\
&\leq 2\|\sqrt{H}\log\left(H\right)'\|_{L^2(\pi)} \|\sqrt{H}\log\left(H\right)\|_{L^2(\pi)} + K\|\sqrt{H}\log\left(H\right)\|_{L^2(\pi)}^2
\end{aligned}$$

the last inequality holds due to Cauchy-Schwarz inequality for the first term, for the second term see that $H' = H\log\left(H\right)'$, using the hypothesis for the derivative of $\log\left(H\right)$, we have that $\| \log (H)' \|_{L^\infty(\pi)} < 1$ and for this part of the proof, we denote $\| \log (H)' \|_{L^\infty(\pi)}$ by K. The above inequality is equivalent to

$$\| \sqrt{H} \log (H) \|_{L^2(\pi)} \leq \frac{2}{1-K} \| \sqrt{H} \log (H)' \|_{L^2(\pi)}$$

Therefore, we obtain that

$$Ent_\pi (H) \leq \frac{2}{1-K} \| \sqrt{H} \log (H)' \|^2_{L^2(\pi)}$$

Using that $H(x) = h(x)/h(0)$, we have that

$$Ent_\pi (h) \leq \frac{2}{1-K} \int h(\log (h)')^2 \pi(dx)$$

For the next step, we consider $\pi_{0,I}$ and apply Proposition 2.2 in LEDOUX (1997). If we further assume that $K = \| \nabla \log (h) \|_\infty = \max_{1 \leq i \leq d} \| \partial_i \log (h) \|_\infty < 1$, we have that

$$Ent_{\pi_{0,I}} (h) \leq \sum_{i=1}^{d} \mathbb{E}_{\pi_{0,I}} (Ent_{\pi_i} (h)) \leq \frac{2}{1-K} \int h \| \nabla \log (h) \|^2_2 \pi(dx)$$

For the last step, we extend the last inequality which is obtained for $\pi_{0,I}$ to the probability measure $\pi_{m,B}$, applying a change of variable in the same way as we did in the previous proof. Employing the affine map $g(x) := Bx + m$ and assuming that $K = \| \nabla \log (h \circ g) \|_\infty < 1$, we get that

$$\begin{aligned} Ent_{\pi_{m,B}} (h) = Ent_{\pi_{0,I}} (h \circ g) &\leq \frac{2}{1-K} \int h \circ g \| \nabla \log (h \circ g) \|^2_2 \, d\pi_{0,I} \\ &= \frac{2}{1-K} \int h \| B^T \nabla \log (h) \|^2_2 \, d\pi_{m,B} \end{aligned}$$

Let us denote by b^i the columns of B, i.e. $B = \left(b^1, \dots b^d\right)$. Then using the chain rule, we get that K can be written as follows,

$$K = \max \left\{ \| \left(b^i, \nabla \log (h)\right) \|_\infty : i \in \{1, \dots, d\} \right\}$$

Now, let us consider the vector infinite-norm which is defined on $\mathbb{R}^d$ and denote it by $\| \cdot \|_{\infty,d}$. Next, we define its dual norm $\| \cdot \|^*_{\infty,d}$ through the following expression $\|b\|^*_{\infty,d} = \sup_{\|x\|_{\infty,d}=1} | (b, x) |$. Also, notice that $\| \cdot \|^*_{\infty,d} = \| \cdot \|_{1,d}$, thus we have that

$$\| \left(b^i, \nabla \log (h)\right) \|_\infty \leq \|b^i\|^*_\infty \| \nabla \log (h) \|_\infty = \|b^i\|_{1,d} \| \nabla \log (h) \|_\infty$$

See for example that the induced matrix norm by the vector norm $\| \cdot \|^*_{\infty,d}$ of the matrix B^T is equal to $\|B^T\|^*_\infty = \max \left\{ \|b^i\|_1 : i \in \{1, \dots d\} \right\}$. Thus, we get that

$$K \leq \|B^T\|^*_\infty \| \nabla \log (h) \|_\infty .$$

Proof of Theorem 4.4.3 Let us consider the matrix $U = BP$, where P corresponds to the permutation matrix considered from the definition of $P_r \in \mathcal{P}_{\pi_{m,B},r}$. In addition, we consider the standard basis of $\mathbb{R}^d$ denoted by e_i, i.e. e_i is the d-vector with zero everywhere except in the i-th coordinate which take the value 1, then, we define matrices $C_r := (e_1, \ldots e_r)$ and $C_{d-r} := (e_{r+1}, \ldots e_d)$. Furthermore, we can define the matrices $U_r := UC_r$ and $U_\perp := UC_{d-r}$, see also that "it holds" $U = (U_r|U_\perp)$.

Observe also that the columns of the matrices U_r and $U_\perp$ forms a basis of the subspaces $Im(P_r)$ and $Ker(P_r)$, respectively. Therefore, the matrices U_r and $U_\perp$ satisfies the assumptions stated in Section 4.3.1 and in particular for the construction of the probability measures $\pi_\perp(\cdot|\xi_r)$ and $\pi_r(\cdot)$, as it is stated in the integration rule (4.12).

An advantage of choosing that particular U is that $\pi_\perp(d\xi_\perp|\xi_r)$ is independent of the choice of ξ_r. In order to show our claim, we observe that for every $\xi := \left(\begin{smallmatrix}\xi_r \\ \xi_\perp\end{smallmatrix}\right) \in \mathbb{R}^d$, where $\xi_r \in \mathbb{R}^r$ and $\xi_\perp \in \mathbb{R}^{d-r}$, we have that

$$\|B^{-1}U(\xi - \tilde{m})\|_{1,d} = \|C_r^T PB^{-1}U_r(\xi_r - \tilde{m}_r)\|_{1,r} + \|C_{d-r}^T PB^{-1}U_\perp(\xi_\perp - \tilde{m}_\perp)\|_{1,d-r}$$

where $\tilde{m} := U^{-1}m$, $\tilde{m}_r := C_r^T\,\tilde{m}$, $\tilde{m}_\perp := C_\perp^T\tilde{m}$ and $\|\cdot\|_{1,n}$ is the 1-norm is defined on $\mathbb{R}^n$, i.e. $\|x\|_{1,n} := \sum_{i=1}^n |x_i|$. Note that the last equality, is obtained readily by observing the following matrix

$$P^T B^{-1} U = \left(\begin{array}{c|c} I_r & \mathbf{0} \\ \hline \mathbf{0} & I_{d-r} \end{array}\right) \tag{4.26}$$

where I_r and I_{d-r} are the corresponding identity matrices. Hence, we have that $C_r^T P^T B^{-1}U = (I_r|0)$. Next, using the definition of $\pi_\perp(d\xi_\perp|\xi_r)$ and the expression for the conditional expectation $p_\perp$, see expression (4.11), we get that

$$\pi_\perp(d\xi_\perp|\xi_r) = p_\perp\left(\xi_\perp|U_r\xi_r\right)d\xi_\perp = \frac{e^{-\|C_r^T P^T B^{-1}U_\perp(\xi_\perp - \tilde{m}_\perp)\|_{1,d-r}}}{\int_{\mathbb{R}^{d-r}} e^{-\|C_r^T P^T B^{-1}U_\perp(\xi_\perp - \tilde{m}_\perp)\|_{1,d-r}}\mathrm{d}\xi_\perp'}$$

Therefore the structure of U combined with the assumption on the projection P_r implies that $\pi_\perp(d\xi_\perp|\xi_r)$ is independent of the choice of ξ_r, i.e. $\pi_\perp(d\xi_\perp|\xi_r) = \pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)$.

For the purpose of this proof, the quantity of interest is divided into the following two integrals,

$$\begin{aligned}\int f\log\left(\frac{f}{\mathbb{E}_{\pi_{m,B}}\left(f|\sigma(P_r)\right)}\right)d\pi_{m,B} &= \int_C f\log\left(\frac{f}{\mathbb{E}_{\pi_{m,B}}\left(f|\sigma(P_r)\right)}\right)d\pi_{m,B} \\ &\quad + \int_{C^c} f\log\left(\frac{f}{\mathbb{E}_{\pi_{m,B}}\left(f|\sigma(P_r)\right)}\right)d\pi_{m,B}\end{aligned} \tag{4.27}$$

For the estimation of the first integral, we apply the local version of the modified LSI which is obtained in Theorem 4.4.2. Observe that the probability $\pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)$ admits the assumptions of that Theorem. But first, let us consider the set C and define the set $\tilde{C}^r$ and $\tilde{C}\left(\xi_r\right)$ as follows,

$$\tilde{C}^r := \left\{\xi_r \in \mathbb{R}^r : \exists \xi_\perp,\ U\left(\begin{smallmatrix}\xi_r\\ \xi_\perp\end{smallmatrix}\right) \in C\right\} = C_r^T U^{-1}(C)$$

and for a given $\xi_r \in \tilde{C}^r$

$$\tilde{C}\left(\xi_r\right) := \left\{\xi_\perp \in \mathbb{R}^{d-r} : U\left(\begin{smallmatrix}\xi_r\\ \xi_\perp\end{smallmatrix}\right) \in C\right\}$$

Using the integration rule (4.12) in a similar way as we use it for getting the expression in (4.16), we get that

$$\int_C f \log\left(\frac{f}{\mathbb{E}_{\pi_{m,B}}\left(f|\sigma(P_r)\right)}\right) d\pi_{m,B} \leq \int_{\tilde{C}^r} Ent_{\pi_{\tilde{m}_\perp, I_{d-r}}}\left(1\left(\cdot \in \tilde{C}\left(\xi_r\right)\right) f \circ U\right)\ \pi_r(d\xi_r)$$

for getting the above inequality, we only needs to apply the following inequality, for every $\xi_r \in \tilde{C}^r$, we have that

$$\int_{\tilde{C}(\xi_r)} f \circ U\left(\begin{smallmatrix}\xi_r\\ \xi_\perp\end{smallmatrix}\right) \pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp) \leq \int f \circ U\left(\begin{smallmatrix}\xi_r\\ \xi_\perp\end{smallmatrix}\right) \pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)$$

For simplicity, let us denote $g\left(\xi_\perp\right) = f\left(U_r\xi_r + U_\perp\xi_\perp\right)$ and apply theorem 4.4.2 on $\pi_{\tilde{m}_\perp, I_{d-r}}$, we have that

$$Ent_{\pi_{\tilde{m}_\perp, I_{d-r}}}\left(1\left(\cdot \in \tilde{C}\left(\xi_r\right)\right) g\right) \leq \frac{2}{c_{\xi_r}} \int_{\tilde{C}(\xi_r)} \|\nabla g(\xi_\perp)\|_{2,d-r}^2\ \pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp) + \tilde{M}_{\xi_r}\ \pi_{\tilde{m}_\perp, I_{d-r}}\left(\tilde{C}\left(\xi_r\right)^c\right)$$

where $\tilde{M}_{\xi_r} := a_{\xi_r} M_{\xi_r}$ and c_{ξ_r}, a_{ξ_r} M_{ξ_r} are given in the following forms,

$$c_{\xi_r} = \inf\left\{f \circ U\left(\begin{smallmatrix}\xi_r\\ \xi_\perp\end{smallmatrix}\right) : \xi_\perp \in \overline{\tilde{C}\left(\xi_r\right)}\right\}$$

$$M_{\xi_r} := \sup\left\{f \circ U\left(\begin{smallmatrix}\xi_r\\ \xi_\perp\end{smallmatrix}\right) : \xi_\perp \in \overline{\tilde{C}\left(\xi_r\right)}\right\}$$

$$a_{\xi_r} := \begin{cases} 1, & \text{if } \tilde{C}\left(\xi_r\right) - \tilde{m}_\perp \text{ is cube} \\ d, & \text{otherwise} \end{cases}$$

It is obvious that the definition of cube is invariant under permutations and translations, i.e. if $C \subset \mathbb{R}^d$ is cube, $P \in \mathbb{R}^{d\times d}$ permutation matrix and $m \in \mathbb{R}^d$, we have that $P(C-m)$ is cube. Notice that if $B^{-1}(C)$ is cube, then we have that $U^{-1}(C)$ is cube, so there exist $A \subseteq \mathbb{R}^r$ and $B \subseteq \mathbb{R}^{d-r}$ such that $U^{-1}(C) = A_r \times A_\perp$. See also that by definition of $\tilde{C}^r$ and $\tilde{C}\left(\xi_r\right)$, $U^{-1}(C)$ can be written as follows

$$U^{-1}(C) := \left\{\left(\begin{smallmatrix}\xi_r\\ \xi_\perp\end{smallmatrix}\right) : \xi_r \in \tilde{C}^r, \xi_\perp \in \tilde{C}\left(\xi_r\right)\right\}$$

Therefore, if $B^{-1}(C)$ is cube, we have that for every $\xi_r \in \tilde{C}^r$, the set $\tilde{C}(\xi_r) = A_\perp$, see also that $A_\perp$ is cube by definition. That implies that $\sup\{a_{\xi_r} : \xi_r \in \tilde{C}^r\} \leq a$, where a is defined as follows,

$$a := \begin{cases} 1, & \text{if } B^{-1}(C) \text{ is cube} \\ d, & \text{otherwise} \end{cases}$$

See also that

$$\begin{aligned} \inf\left\{c_{\xi_r} : \xi_r \in \tilde{C}^r\right\} &\geq \inf\left\{f \circ U \begin{pmatrix} \xi_r \\ \xi_\perp \end{pmatrix} : \xi_r \in \tilde{C}^r, \xi_\perp \in \overline{\tilde{C}(\xi_r)}\right\} \\ &\geq \inf\left\{f(x) : x \in \overline{C}\right\} = c \end{aligned}$$

Similar to the lower bound of c_{ξ_r}, we have that

$$\sup\left\{c_{\xi_r} : \xi_r \in \tilde{C}^r\right\} \leq \sup\left\{f(x) : x \in \overline{C}\right\} = M$$

Let us substitute g with $f \circ U$, also see that using the the chain rule, we have that $\nabla g(\xi_\perp) := U_\perp^T \nabla f(U_r\xi_r + U_\perp \xi_\perp)$,

$$\begin{aligned} &Ent_{\pi_{\tilde{m}_\perp, I_{d-r}}}\left(1\left(\cdot \in \tilde{C}(\xi_r)\right) f \circ U\right) \\ &\quad \leq \frac{2}{c} \int_{\tilde{C}(\xi_r)} \|U_\perp^T \nabla f(U_r\xi_r + U_\perp\xi_\perp)\|_{2,d-r}^2 \ \pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp) + \tilde{M} \ \pi_{\tilde{m}_\perp, I_{d-r}}\left(\tilde{C}(\xi_r)^c\right) \end{aligned}$$

Let us integrate the above inequality with respect to $\pi_r(d\xi_r)$ over the set $\tilde{C}^r$,

$$\begin{aligned} &\int_{\tilde{C}^r} Ent_{\pi_{\tilde{m}_\perp, I_{d-r}}}\left(1\left(\cdot \in \tilde{C}(\xi_r)\right) f \circ U\right) \pi_r(d\xi_r) \\ &\quad \leq \frac{2}{c} \int_{\tilde{C}^r} \int_{\tilde{C}(\xi_r)} \|U_\perp^T (\nabla f)(U_r\xi_r + U_\perp\xi_\perp)\|_{2,d-r}^2 \ \pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)\, \pi_r(d\xi_r) \\ &\qquad\qquad + \tilde{M} \int_{\tilde{C}^r} \pi_{\tilde{m}_\perp, I_{d-r}}\left(\tilde{C}(\xi_r)^c\right) \pi_r(d\xi_r) \\ &\quad = \frac{2}{c} \int_C \|U_\perp^T (\nabla f)(x)\|_{2,d-r}^2 \ \pi_{m,B}(dx) + \tilde{M} \ \pi_{m,B}\left(C^c\right) \end{aligned}$$

In the following part of the proof, we estimate the second term of (4.27) which is an integral involving the values of f over the set C^c. By applying the integration rule which is given in (4.12), we obtain the following

$$\begin{aligned} &\int_{C^c} f \log\left(\frac{f}{\mathbb{E}_{\pi_{m,B}}\left(f|\sigma\left(P_r\right)\right)}\right) d\pi_{\mathrm{m,B}} = \\ &\qquad \int \int_{\left(U^{-1}C\right)^c} f \circ U \log\left(\frac{f \circ U}{\mathbb{E}_{\pi_{m,B}}\left(f|\sigma\left(P_r\right)\right) \circ U}\right) \pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)\pi_r(d\xi_r) \end{aligned}$$

It is easy to see that the above integral over $\left(U^{-1}C\right)^c$, it is bounded by the following two integrals,

$$\begin{aligned} &\int_{\tilde{C}^r} \int_{\tilde{C}(\xi_r)^c} f \circ U \log\left(\frac{f \circ U}{\int f \circ U d\pi_{\tilde{m}_\perp, I_{d-r}}}\right) \pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)\pi_r(d\xi_r) + \\ &\qquad\qquad \int_{\left(\tilde{C}^r\right)^c} Ent_{\pi_{\tilde{m}_\perp, G}}\left(f \circ U\right) \pi_r(d\xi_r) \end{aligned} \tag{4.28}$$

Before, we start with the estimation of the last two integrals, we use the log-Lipschitz condition of f to obtain a bound for the Logarithm that appears in the above two integrals. Since f is log-Lipschitz, we have that $\|\nabla \log(f)\|_{L^{\infty}(\pi_{m,B}),\infty}$ is finite. In addition, the definition of log-Lipschitz for positive functions implies that for every $x, y \in \mathbb{R}^d$, it holds

$$f(y)e^{-K\|x-y\|_1} \leq f(x) \leq f(y)e^{K\|x-y\|_1}$$

where $K = \|\nabla \log(f)\|_{L^{\infty}(\pi_{m,B}),\infty}$.

Let us also consider the following notation, for every $\xi_\perp, \xi'_\perp \in \mathbb{R}^{d-r}$ and $\xi_r \in \mathbb{R}^r$, we denote with $\xi, \xi' \in \mathbb{R}^d$ the following two vectors $\begin{pmatrix} \xi_r \\ \xi_\perp \end{pmatrix}$, $\begin{pmatrix} \xi_r \\ \xi'_\perp \end{pmatrix}$, respectively. Then, we have that for every $\xi_r \in \mathbb{R}^r$ and $\xi_\perp \in \mathbb{R}^{d-r}$, it holds

$$\begin{aligned}\int \frac{f \circ U(\xi')}{f \circ U(\xi)} \pi_{\tilde{m}_\perp,G}(d\xi'_\perp) &\geq \int e^{-K\|U_\perp(\xi_\perp-\xi'_\perp)\|_1} \pi_{\tilde{m}_\perp,G}(d\xi'_\perp) \\ &\geq \int e^{-K\|U_\perp\|_1^*\|\xi_\perp-\xi'_\perp\|_1} \pi_{\tilde{m}_\perp,G}(d\xi'_\perp)\end{aligned}$$

where $\|U_\perp\|_1^*$ is the dual-norm of $U_\perp$ with respect to $\|\cdot\|_1$, i.e. $\|U_\perp\|_1^* = \sup_{\xi_\perp \in \mathbb{R}^{d-r}} \frac{\|U_\perp \xi_\perp\|_{1,d}}{\|\xi_\perp\|_{1,d-r}}$. Then, using the triangular inequality on the logarithm of (4.28), we have that

$$\begin{aligned}\log\left(\frac{f \circ U}{\int f \circ U d\pi_{\tilde{m}_\perp,I_{d-r}}}\right) &\leq K\|U_\perp\|_1^*\|\xi_\perp - \tilde{m}_\perp\|_1 + \log\left(\frac{1}{\int e^{-K\|U_\perp\|_1^*\|\xi'_\perp-\tilde{m}_\perp\|_1}\pi_{\tilde{m}_\perp,G}(d\xi'_\perp)}\right) \\ &= K\|U_\perp\|_1^*\|\xi_\perp - \tilde{m}_\perp\|_1 + (d-r)\log(K\|U_\perp\|_1^*+1)\end{aligned} \tag{4.29}$$

the last equality holds, since for any positive α we have

$$\int_{\mathbb{R}^{d-r}} e^{-\alpha\|\tilde{\xi}_\perp-\tilde{m}_\perp\|_{1,d-r}} \pi_{\tilde{m}_\perp,I_{d-r}}(d\xi_\perp) = \frac{1}{(\alpha+1)^{d-r}}$$

Next, by applying the inequality (4.29) on the second integral of (4.28), we have that

$$\begin{aligned}&\int_{(\tilde{C}^r)^c} Ent_{\xi_\perp\sim\pi_{\tilde{m}_\perp,G}}(f \circ U(\xi))\,\pi_r(d\xi_r) \\ &= \int_{(\tilde{C}^r)^c}\int_{\mathbb{R}^{d-r}} f \circ U(\xi)\log\left(\frac{f \circ U(\xi)}{\int f \circ U(\xi')\pi_{\tilde{m}_\perp,I_{d-r}}(d\xi'_\perp)}\right)\pi_{\tilde{m}_\perp,I_{d-r}}(d\xi_\perp)\pi_r(d\xi_r) \\ &\leq M_{C^c}\int_{(\tilde{C}^r)^c}\int_{\mathbb{R}^{d-r}} K\|U_\perp\|_1^*\|\xi_\perp - \tilde{m}_\perp\|_1 + (d-r)\log(K\|U_\perp\|_1^*+1)\,\pi_{\tilde{m}_\perp,I_{d-r}}(d\xi_\perp)\pi_r(d\xi_r) \\ &= M_{C^c}(d-r)(K\|U_\perp\|_1^* + \log(K\|U_\perp\|_1^*+1))\int_{(\tilde{C}^r)^c}\int_{\mathbb{R}^{d-r}}\pi_{\tilde{m}_\perp,I_{d-r}}(d\xi_\perp)\pi_r(d\xi_r)\end{aligned} \tag{4.30}$$

where $M_{C^c} = \sup\{f(x) : x \in C^c\}$ and the last equality holds, since we have that

$$\int_{\mathbb{R}^{d-r}} \|\tilde{\xi}_\perp - \tilde{m}_\perp\|_{1,d-r}\pi_{\tilde{m}_\perp,I_{d-r}}(d\xi_\perp) = d-r \tag{4.31}$$

For the estimation of the first integral of (4.28), we consider Λ similar to the one is assumed in the theorem and then we define sets $\tilde{\Lambda}^r$, $\tilde{\Lambda}(\xi_r)$ similar to the definition of $\tilde{C}^r$, $\tilde{C}(\xi_r)$.

Let us consider the case where $\xi_r \in \tilde{C}^r \setminus \tilde{\Lambda}^r$. Applying the inequality (4.29) on the following integral, we get that

$$\int_{\tilde{C}(\xi_r)^c} f\circ U \log\left(\frac{f\circ U}{\int f\circ U d\pi_{\tilde{m}_\perp, I_{d-r}}}\right)\pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)$$
$$\leq M_{C^c}\int_{\tilde{C}(\xi_r)^c} K\|U_\perp\|_1^*\,\|\xi_\perp - \tilde{m}_\perp\|_1 + (d-r)\log\left(K\|U_\perp\|_1^* + 1\right)\pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)$$

Notice that

$$\left\{\begin{pmatrix}\xi_r\\ \xi_\perp\end{pmatrix} : \xi_r \in \tilde{C}^r\setminus\tilde{\Lambda}^r, \xi_\perp \in \tilde{C}\left(\xi_r\right)\right\} \subseteq U^{-1}(C\setminus\Lambda)$$

Then, it is obvious that for every $\xi_r \in \tilde{C}^r$, we have that $\tilde{C}\left(\xi_r\right) \subseteq C_{d-r}^T U^{-1}(C\setminus\Lambda)$. Thus, the following holds

$$\sup\left\{\pi_{\tilde{m}_\perp, I_{d-r}}(\tilde{C}(\xi_r)) : \xi_r \in \tilde{C}^r\setminus\tilde{\Lambda}^r\right\} \leq \pi_{\tilde{m}_\perp, I_{d-r}}(C_{d-r}^T U^{-1}(C\setminus\Lambda)\,)$$
$$= \pi_{0, I_{d-r}}(C_{d-r}^T U^{-1}(C\setminus\Lambda\, - m)\,) =: \beta$$

which implies that

$$\sup_{\xi_r\in\tilde{C}^r\setminus\tilde{\Lambda}^r}\frac{\pi_{\tilde{m}_\perp, I_{d-r}}(\tilde{C}(\xi_r))}{\pi_{\tilde{m}_\perp, I_{d-r}}(\tilde{C}(\xi_r)^c)} \leq \frac{\beta}{1-\beta}$$

Combining the above inequality with the integral of (4.31), we obtain that

$$\int_{\tilde{C}(\xi_r)^c}\|\tilde{\xi}_\perp - \tilde{m}_\perp\|_{1,d-r}\pi_{\tilde{m}_\perp, I_{d-r}}\left(d\xi_\perp\right) \leq \frac{d-r}{1-\beta}\,\pi_{\tilde{m}_\perp, I_{d-r}}\left(\tilde{C}\left(\xi_r\right)^c\right)$$

Hence, integrating with respect to $\pi_r(d\xi_r)$ over the set $\tilde{C}^r\setminus\tilde{\Lambda}^r$, we get that

$$\int_{\tilde{C}^r\setminus\tilde{\Lambda}^r}\int_{\tilde{C}(\xi_r)^c} f\circ U\log\left(\frac{f\circ U}{\int f\circ U d\pi_{\tilde{m}_\perp, I_{d-r}}}\right)\pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)\pi_r(d\xi_r)$$
$$\leq M_{C^c}\left(d-r\right)\left(\frac{K\|U_\perp\|_1^*}{1-\beta} + \log\left(K\|U_\perp\|_1^* + 1\right)\right)\int_{\tilde{C}^r\setminus\tilde{\Lambda}^r}\int_{\tilde{C}(\xi_r)^c}\pi_{\tilde{m}_\perp, G}(d\xi_\perp)\pi_r(d\xi_r) \tag{4.32}$$

Next, we consider the case where $\xi_r \in \tilde{\Lambda}^r$ and define the following constant

$$\tilde{\lambda} := \inf\left\{\pi_{\tilde{m}_\perp, I_{d-r}}\left(\tilde{\Lambda}\left(\xi_r\right)\right) : \xi_r \in \tilde{\Lambda}^r\right\} = \inf\left\{\pi_{0, I_{d-r}}\left(\tilde{\Lambda}\left(\xi_r\right) - \tilde{m}_\perp\right) : \xi_r \in \tilde{\Lambda}^r\right\}$$

Using the definition of $\tilde{\Lambda}\left(\xi_r\right)$ and U, we have that

$$\pi_{0, I_{d-r}}\left(\tilde{\Lambda}\left(\xi_r\right) - \tilde{m}_\perp\right) = \pi_{0, I_{d-r}}\left(\left\{\xi_\perp - \tilde{m}_\perp \in \mathbb{R}^{d-r} : \begin{pmatrix}\xi_r - \tilde{m}_r\\ \xi_\perp - \tilde{m}_\perp\end{pmatrix} \in P^T B^{-1}\left(\Lambda - m\right)\right\}\right)$$
$$\geq \min\left\{\pi_{0,1}\left(e_i^T B^{-1}(\Lambda - m)\right) : e_i \text{ is the standard basis of } \mathbb{R}^d\right\}^{d-r}$$

for the last inequality, we only need to use that $\pi_{0, I_{d-r}}$ is the product of $\pi_{0,1}$ over *span* $\{e_1, \dots e_d\}$. Next, using that $B^{-1}(\Lambda)$ is cube and P is the permutation matrix, we get that $U^{-1}(\Lambda - m) = PB^{-1}(\Lambda - m)$ is cube. Thus, it is obvious that for every e_i,

$\pi_{0,1}\left(e_i^T B^{-1}(\Lambda - m)\right)$ is positive, and so, it follows that $\tilde{\lambda} > 0$. Then, the first term of (4.28) can be bounded as follows

$$\int_{\tilde{C}(\xi_r)^c} f\circ U \log\left(\frac{f\circ U}{\int f\circ U d\pi_{\tilde{m}_\perp, I_{d-r}}}\right)\pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)$$
$$\leq \int_{\tilde{C}(\xi_r)^c} f\circ U \log\left(\frac{f\circ U}{\int_{\tilde{\Lambda}(\xi_r)} f\circ U d\pi_{\tilde{m}_\perp, I_{d-r}}}\right)\pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)$$
$$\leq \int_{\tilde{C}(\xi_r)^c} f\circ U \log\left(\frac{M_{C^c}}{c\tilde{\lambda}}\right)\pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)$$

Let us integrate the above inequality with respect to $\pi_r(d\xi_r)$ over the set $\tilde{\Lambda}^r$,

$$\begin{aligned}\int_{\tilde{\Lambda}^r}\int_{\tilde{C}(\xi_r)^c} f\circ U \log\left(\frac{f\circ U}{\int f\circ U d\pi_{\tilde{m}_\perp, I_{d-r}}}\right)\pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)\pi_r(d\xi_r) \leq \\ M_{C^c}\max\left\{\log\left(\frac{M_{C^c}}{c\,\tilde{\lambda}}\right), 0\right\}\int_{\tilde{\Lambda}^r}\int_{\tilde{C}(\xi_r)^c}\pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)\pi_r(d\xi_r)\end{aligned} \tag{4.33}$$

Applying inequalities (4.30), (4.32) and (4.33) on the relation (4.28), we obtain that

$$\int_{C^c} f\log\left(\frac{f}{\mathbb{E}_{\pi_{m,B}}\left(f|\sigma\left(P_r\right)\right)}\right)d\pi_{\mathrm{m,B}}$$
$$\leq M_{C^c}\max\left\{(d-r)\left(\frac{K\|U_\perp\|_1^*}{1-\beta} + \log\left(K\|U_\perp\|_1^* + 1\right)\right), \log\left(\frac{M_{C^c}}{c\,\tilde{\lambda}}\right), 0\right\}\pi_{m,B}(C^c)$$

Therefore, we have that

$$\int f\log\left(\frac{f}{\mathbb{E}_{\pi_{m,B}}\left(h|\sigma\left(P_r\right)\right)}\right)\ \pi_{m,B}(dx) \leq \frac{2}{c}\int_C \|U_\perp^T\nabla f(x)\|_{2,d-r}^2\ \pi_{m,B}(dx)$$
$$+\max\left\{\tilde{M}, M_{C^c}(d-r)\left(\frac{K\|U_\perp\|_1^*}{1-\beta} + \log\left(K\|U_\perp\|_1^* + 1\right)\right), M_{C^c}\log\left(\frac{M_{C^c}}{c\,\tilde{\lambda}}\right)\right\}\ \pi_{m,B}\left(C^c\right)$$

This complete proves the first two inequalities of the theorem.

The last inequality is a consequence of Theorem 4.4.1. So, proceeding similarly as above, we fix $\xi_r \in \mathbb{R}^r$ and consider the function $g\left(\xi_\perp\right) = f\left(U_r\xi_r + U_\perp\xi_\perp\right)$. Next, using the chain rule, we have that $\nabla\log\left(g(\xi_\perp)\right) = U_\perp^T\nabla\log\left(f(x)\right)|_{x=U_r\xi_r+U_\perp\xi_\perp}$. In order to apply Theorem 4.4.1, we need to define $K_{\xi_r} = \|\nabla\log\left(g\right)\|_{L^\infty(\pi_{\tilde{m}_\perp, I_{d-r}}),\infty}$ for every $\xi_r \in \mathbb{R}^r$. See that for every $\xi_r \in \mathbb{R}^r$, the following holds

$$K_{\xi_r} \leq \|U_\perp^T\nabla\log\left(f\right)\|_{L^\infty(\pi_{m,B}),\infty} \leq K := \|U_\perp^T\|_\infty^*\,\|\nabla\log\left(f\right)\|_{L^\infty(\pi_{m,B}),\infty} < 1$$

Since $K_{\xi_r} < 1$, Theorem 4.4.1 implies that

$$Ent_{\pi_{\tilde{m}_\perp, I_{d-r}}}\left(g\right) \leq \frac{2}{1-K_{\xi_r}}\int g(\xi_\perp)\|\nabla g(\xi_\perp)\|_{2,d-r}^2\pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)$$

Then, we substitute g and ∇g integrate with respect to $\pi_r(d\xi_r)$,

$$\int Ent_{\pi_{\tilde{m}_\perp, I_{d-r}}}\left(f\circ U\right)\pi_r(d\xi_r)$$
$$\leq \frac{2}{1-K}\int\int f\circ U\begin{pmatrix}\xi_r\\ \xi_\perp\end{pmatrix}\|U_\perp^T\left(\nabla\log\left(f\right)\right)\circ U\begin{pmatrix}\xi_r\\ \xi_\perp\end{pmatrix}\|_{2,d-r}^2\ \pi_{\tilde{m}_\perp, I_{d-r}}(d\xi_\perp)\pi_r(d\xi_r)$$

Therefore, using the integration rule (4.12), we get that

$$\int f \log\left(\frac{f}{\mathbb{E}_{\pi_{m,B}}\left(f|\sigma\left(P_r\right)\right)}\right) d\pi_{\mathrm{m,B}} \leq \frac{2}{1-K}\int f \left\|U_\perp^T \nabla \log\left(f\right)\right\|_{2,d-r}^2 \, d\pi_{m,B}$$

Proof of Theorem 4.4.4 First, we know that for any $a \in \mathbb{R}^d$, the 2-norm can be written as follows

$$\|a\|_{2,d}^2 = (a,a) = a^T a = tr(aa^T)$$

Hence, if we apply the second bound coming from Theorem 4.4.3 in equation (4.15), we have that

$$\begin{aligned} D_{KL}\left(\mu||\tilde{\mu}_r\right) &\leq \frac{2}{1-K}\int \left\|U_\perp^T \nabla \log\left(f\right)\right\|_{2,d-r}^2 \frac{f}{Z_f} \, d\pi_{m,B} \\ &= \frac{2}{1-K}\int tr(U_\perp^T \nabla \log\left(f\right)\left(\nabla \log\left(f\right)\right)^T U_\perp) \, d\mu = tr(U_\perp^T H U_\perp) \end{aligned}$$

where $H = \frac{2}{1-K}\int \nabla \log\left(f\right)\left(\nabla \log\left(f\right)\right)^T d\mu \in \mathbb{R}^{d\times d}$.

Also, by the definition of $\overline{\mathcal{P}}_{\pi_{m,B},r}$, we have that for every $P_r \in \overline{\mathcal{P}}_{\pi_{m,B},r}$ there is a corresponding permutation matrix P. Notice also that the structure of a permutation matrix is essentially an rearrangement of the columns of the identity matrix. Hence, for every P there exists a bijective function $\sigma_P : \{1,\dots d\} \to \{1,\dots d\}$ such that $P = \left(e_{\sigma_P(1)},\dots,e_{\sigma_P(d)}\right)$. Then if we denote by $Q = P\left(e_{r+1},\dots,e_d\right) \in \mathbb{R}^{d\times(d-r)}$, we have that $Q = \left(e_{\sigma_P(r+1)},\dots,e_{\sigma_P(d)}\right)$. Recalling the definition of $U_\perp = BP\left(e_{r+1},\dots,e_d\right) = BQ$. Hence, the above integral can be written as follows,

$$tr(U_\perp^T H U_\perp) = tr(Q^T B^T H B Q) = tr(Q^T A Q)$$

where $A := B^T H B$.

Similarly for the first bound, we have that

$$D_{KL}\left(\mu||\tilde{\mu}_r\right) \leq tr(Q^T A Q) + R_{f,C,\pi_{m,B}}$$

where $A = B^T H_C B$ with $H_C := \frac{2}{Z_f\, c}\int_C \nabla f (\nabla f)^T d\pi_{m,B} \in \mathbb{R}^{d\times d}$.

Proof of Proposition 4.4.1 We recall that for every P, there exists a bijective function $\sigma_P : \{1,\dots d\} \to \{1,\dots d\}$ such that $P = \left(e_{\sigma_P(1)},\dots,e_{\sigma_P(d)}\right)$. Thus, $Q = P\left(e_{r+1},\dots,e_d\right) = \left(e_{\sigma_P(r+1)},\dots,e_{\sigma_P(d)}\right)$ and

$$Q^T A Q = \begin{pmatrix} e_{\sigma_P(r+1)}^T \\ \vdots \\ e_{\sigma_P(d)}^T \end{pmatrix} A\left(e_{\sigma_P(r+1)},\dots,e_{\sigma_P(d)}\right) = \left(a_{\sigma_P(r+i),\sigma_P(r+j)}\right)_{i,j\in\{1,\dots(d-r)\}}$$

Hence,

$$tr(Q^T AQ) = \sum_{i=1}^{d-r} a_{\sigma_P(r+i),\sigma_P(r+j)}$$

Then, the minimum comes naturally from the definition of the function τ_A.

Notice that the sequence of y_i in Proposition 4.4.1 is defined according to Gram–Schmidt process which ensures us that $\{y_1, \ldots, y_{d-r}\}$ is orthonormal basis of $Im(B\left(e_{\tau_A(1)}, \ldots, e_{\tau_A(d-r)}\right))$. According to the context of Theorem 4.4.3, there is a projection $P_r \in \overline{\mathcal{P}}_{\pi_{m,B},r}$ such that $U_\perp = B\left(e_{\tau_A(1)}, \ldots, e_{\tau_A(d-r)}\right)$. In Section 4.3.1, we see that $Im(U_\perp) = ker(P_r) = Im(I - P_r)$, it is also obvious that if P_r is a projection then $I - P_r$ is projection too. Therefore, it is sufficient to prove that for any orthonormal sub-basis $\{a_1, \ldots, a_{d-r}\}$ of $\mathbb{R}^d$, the matrix $\sum_{i=1}^{d-r} a_i a_i^T$ is an r-rank projection with $Im\left(\sum\limits_{i=1}^{d-r} a_i a_i^T\right) = \bigoplus\limits_{i=1}^{d-r} \langle a_i \rangle$.

Notice that

$$\begin{aligned}
\left(\sum_{i=1}^{d-r} a_i\, a_i^T\right)^2 &= \sum_{i=1}^{d-r} (a_i\, a_i^T)^2 + \sum_{\substack{i \neq j \\ 1 \leq i,j \leq (d-r)}} a_i\, a_i^T\, a_j\, a_j^T \\
&= \sum_{i=1}^{d-r} (a_i, a_i) a_i\, a_i^T + \sum_{\substack{i \neq j \\ 1 \leq i,j \leq (d-r)}} (a_i, a_j) a_i\, a_j^T = \sum_{i=1}^{d-r} a_i\, a_i^T
\end{aligned}$$

In addition, we observe that $a_i \neq 0$, for every $i \in \{1, \ldots d\}$, so we have that $Ker(a_i\, a_i^T) = \left\{x \in \mathbb{R}^d : a_i\, a_i^T x = 0\right\} = \left\{x \in \mathbb{R}^d : (a_i, x) = 0\right\}$, thus $a_i\, a_i^T$ is 1-rank matrix. Furthermore, we have that $a_j \in Ker(a_i\, a_i^T)$, for every $j \in \{1, \ldots d\} \setminus \{i\}$, thus it holds that $Im\left(\sum\limits_{i=1}^{d-r} a_i a_i^T\right) = \bigoplus\limits_{i=1}^{d-r} \langle a_i \rangle$.

4.5.1 Lemmas

Lemma 4.5.1. *Suppose an interval I and $q = \operatorname{argmin}_{x \in \overline{I}} |x|$. Then, the following inequality holds for any $h \in \mathcal{I}(\pi_{0,1})$ with $h(q) = 0$,*

$$\int_I h \; d\pi_{0,1} \leq \int_I sgn\,(x)\, h'\,(x)\, \pi_{0,1}\,(dx)$$

Proof 4.5.1 (Lemma 4.5.1)**.** *For simplicity, we consider the interval $I = (a, b)$ where $a, b \in \mathbb{R} \cup \{\pm\infty\}$. Next, let us apply the integration by part formula on the left hand side of the above inequality and denote the boundary term by $\mathcal{A}_{\pi_{0,1}} : \mathcal{L}^1 \times \mathbb{R}^2 \to \mathbb{R}$,*

$$\int_a^b h \; d\pi_{0,1} = \mathcal{A}_{\pi_{0,1}}\,(h, a, b) + \int_a^b sgn\,(x)\, h'\,(x)\, \pi_{0,1}\,(dx)$$

The explicit formula of $\mathcal{A}_{\pi_{0,1}}(h,a,b)$ *is given as follows*

$$\mathcal{A}_{\pi_{0,1}}(h,a,b) = \begin{cases} \frac{h(b)e^{b} - h(a)e^{a}}{2}, & \textit{if } b \leq 0 \\ \frac{h(a)e^{-a} - h(b)e^{-b}}{2}, & \textit{if } 0 \leq a \\ h(0) - \frac{h(b)e^{-b} + h(a)e^{a}}{2}, & \textit{if } 0 \in (a,b) \end{cases}$$

By definition $q \in \overline{I}$ *and takes the closure value to* 0*, we also have that* h *is non-negative and* $h(q) = 0$*, so it is straightforward to see that* $\mathcal{A}_{\pi}(h,a,b) \leq 0$*. Therefore, we have that*

$$\int_a^b h \; d\pi \leq \int_a^b sgn(x)\, h'(x)\, \pi(dx) \tag{4.34}$$

Chapter 5

SDEs with vanishing diffusions on bounded intervals

In this chapter we consider the inverse problem of recovering the diffusion and drift functions of a stochastic differential equation from discrete measurements of its solution. It is known that applying the Bayesian approach to this problem gives rise to a well-posed posterior measure provided that the diffusion and drift functions are Hölder continuous Croix et al. (2020). Motivated by applications of this problem in estimating the parameters of birth-and-death processes in a large population, we study the case where the diffusion coefficient depends on a small parameter. We use random perturbation methods to approximate the solution of the stochastic differential equation, and study the convergence properties of this approximation at the limit of vanishing diffusion coefficient. We also formulate and study the resulting approximated posterior measure.

5.1 The inverse problem

Similarly to Section 3.2, we consider the stochastic process $X(t)$ defined on a bounded subinterval of $\mathbb{R}$, denoted by I, such that

$$dX(t) = (a(X(t)) - b(X(t)))\, dt + \frac{1}{\sqrt{N}}\sqrt{a(X(t)) + b(X(t))}dW_t, \quad \forall t \geq t_0 \tag{5.1}$$

with initial value $X(t_0) = x_0$. Furthermore, we assume that N is a large number, i.e. $N \gg 1$. Also, we have that the drift and diffusion coefficients are in the form of $a(x) - b(x)$ and $\sqrt{a(x) + b(x)}$, respectively. For simplicity, let us denote them as follows:

$$\mu(x) := a(x) - b(x) \text{ and } \sigma(x) := \sqrt{a(x) + b(x)}. \tag{5.2}$$

In addition, $(W_t)_{t\geq 0}$ is the standard Brownian motion defined over a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$. Let us assume that $I = [0,1]$ and the behaviour of X on the boundary is described as follows, the left bound is an absorbing barrier and the right bound is a reflective barrier. For clarity in what follows, we denote the solution of the above equation X at time t for a given a and b by $X^{(a,b)}(t)$.

Then, we consider the problem of recovering a and b using noisy discrete measurements of a realisation of $X^{(a,b)}(t)$. Thus, assuming a given time discretisation of the interval $[0,T]$ denoted by $0 < t_1 < \ldots t_J \leq T$, the inverse problem is given as follows: estimate (a,b) given

$$y_i = X^{(a,b)}(t_i) + \epsilon\eta_i, \qquad \text{for every } i \in \{1, \ldots J\} \tag{5.3}$$

where $\epsilon > 0$, and η_i are i.i.d. with $\eta_i \sim N(0,1)$. As we mention above, we denote drift and diffusion coefficients by μ and σ, respectively. If $a+b$ is positive, we observe that there is invertible matrix that maps $(a,b) \to (\mu,\sigma)$. That is, we can restate the inverse problem as follows: estimate μ and σ given

$$y_i = X^{(\mu,\sigma)}(t_i) + \epsilon\eta_i, \qquad \text{for every } i \in \{1, \ldots J\} \tag{5.4}$$

where $\epsilon > 0$, and η_i are i.i.d. with $\eta_i \sim N(0,1)$.

Let us denote by π_0 and π_N^y the prior and the posterior distribution of the above inverse problem. We note that the index N of the measure π_N^y may be omitted at times.

Now, we are interested in evaluating the likelihood of the problem (i.e. the unnormalised Radon-Nikodym of π^y with respect to π_0) in the case of no measurement noise, i.e. $\epsilon = 0$ (see the remark below for the noisy case). We note that $X^{(\mu,\sigma)}(\cdot)$ is an Ito diffusion, thus it satisfies the Markov property. In addition, let us assume that the transition probability of $X^{(\mu,\sigma)}(\cdot)$ admits a density function. Notice also that Bayes' formula implies that the likelihood is given as the density function of Y (the distribution of the data) for fixed (μ,σ). Combining the last two observation with the assumption about the density of $X^{(\mu,\sigma)}(\cdot)$, we have that the likelihood is given as the product of the transition density function of $X^{(\mu,\sigma)}(\cdot)$ evaluated in points (y_i, t_i). Thus, the likelihood $\mathcal{L}^{y,s}(\mu,\sigma)$ can be written as follows

$$\frac{\mathrm{d}\pi^y}{\mathrm{d}\pi_0}(\mu,\sigma) \propto \mathcal{L}^{y,s}(\mu,\sigma) := \prod_{i=0}^{J-1} p^{(\mu,\sigma)}(y_{i+1}, t_{i+1}; y_i, t_i) \tag{5.5}$$

According to Section 3.3 and since $X^{(\mu,\sigma)}(\cdot)$ admits a transition probability density function, $p^{(\mu,\sigma)}(t,y;s,x)$, for $t \geq s$, satisfies the associated backward Kolmogorov equation

$$-\frac{\partial\, p^{(\mu,\sigma)}(s,x)}{\partial s} = \frac{\sigma^2(x)}{2N}\,\frac{\partial^2\, p^{(\mu,\sigma)}(s,x)}{\partial x^2} + \mu(x)\,\frac{\partial\, p^{(\mu,\sigma)}(s,x)}{\partial x}$$
$$\lim_{s\nearrow t} p^{(\mu,\sigma)}(t,y;s,x) = \delta(y-x) \tag{5.6}$$

variables (y,t) are omitted in the above expression, since these are fixed.

As we can see in Croix et al. (2020), in the case of noisy measurements, i.e. $\epsilon > 0$, we have that the likelihood of the problem is similar to the above one where the transition probability density $p^{(\mu,\sigma)}$ is replaced by the following

$$\tilde{p}^{(\mu,\sigma)}(\tilde{y}_{i+1},t_{i+1};\tilde{y}_i,t_i) = \int\int p^{(\mu,\sigma)}(z,t_{i+1};w,t_i) f_{N(0,\epsilon^2)}(\tilde{y}_{i+1}-z) f_{N(0,\epsilon^2)}(\tilde{y}_i - w)dzdw \tag{5.7}$$

since the noise term has a Gaussian distribution, we denote by $f_{N(0,\epsilon^2)}$ the density function of $N(0,\epsilon^2)$. For simplicity, let us define the following notation: we define convolution at time t, for given function f convoluted with transition density function p as follows

$$p *_t f\,(y,t;x,s) := \int_{\mathbb{R}} p\,(z,t;x,s)\; f\,(y-z)\; dz$$

Also, we define convolution at time s, for given function f convoluted with transition density function p as follows

$$p *_s f\,(y,t;x,s) := \int_{\mathbb{R}} p\,(y,t;z,s)\; f\,(x-z)\; dz$$

Then $\tilde{p}$ can be rewritten in the following way

$$\begin{aligned}\tilde{p}(y_{i+1},t_{i+1};x_i,t_i) :&= \int_{\mathbb{R}} p *_{t_{i+1}} f_{N(0,\epsilon^2)}\,(y_{i+1},t_{i+1};q,t_i)\, f_{N(0,\epsilon^2)}(y_i - q)dq \\ &= \left(p *_{t_{i+1}} f_{N(0,\epsilon^2)}\right) *_{t_i} f_{N(0,\epsilon^2)}\,(y_{i+1},t_{i+1};y_i,t_i)\end{aligned}$$

Let us now restate the appropriate framework for the well-posedness of the Bayesian approach for the current inverse problem. This framework has been introduced in Croix et al. (2020). Therein, they define the following Banach space,

$$\Lambda = \left\{(\mu,\sigma) : \mu,\sigma^2 \in C^{0,1}([0,1]),\; m_\sigma := \min_{x\in[0,1]} \sigma(x) > 0\right\}$$

endowed with the following norm: given $(\mu,\sigma^2) \in C^{0,1}([0,1])^2$, we define

$$\|(\mu,\sigma^2)\| := \|\mu\|_{0,1} + \|\sigma^2\|_{0,1}$$

where $\|\cdot\|_{0,1}$ is 1-Hölder norm. In general, we define α-Hölder norm for some $\alpha \in (0,1]$ as follows: given $f \in C^{0,\alpha}([0,1])$

$$\|f\|_{0,\alpha} := \|f\|_{L^\infty([0,1])} + \sup_{x\in[0,1]} \frac{\|f(x)-f(y)\|}{\|x-y\|^\alpha}.$$

Let us then restate the following theorem which provides us with the well-definedness of the posterior emerging from the above problem, and also, the stability of the Bayesian approach. Its proof can be found in Croix et al. (2020).

Theorem 5.1.1. *Let T>0, $J \geq 1$ and consider $y = (y_1, \ldots, y_J) \in (0,1)^n$ to be the data at times $t = (t_1, \ldots, t_J)$ with $0 < t_1 \leq \cdots \leq t_J \leq T$. Suppose that π_0 a probability measure with $\pi_0(\Lambda) = 1$ and there exists $C > 0$ and $q > 2$ such that*

$$\mathbb{E}^{\pi_0}\left(\exp\left(C m_\sigma^{(1-q)} \|(\mu, \sigma)\|^q \right) \right) < \infty$$

Then there exists a unique posterior measure $\pi^{y,s}$ given by

$$\frac{d\pi^{y,s}}{d\pi_0}(\mu, \sigma) = \frac{1}{Z(y,s)} \mathcal{L}^{y,s}(\mu, \sigma)$$

where $Z(y, s)$ is normalisation constant which defined as follows

$$Z(y,s) = \int_\Lambda \mathcal{L}^{y,s}(\mu, \sigma)\, d\pi_0(\mu, \sigma) \in (0, \infty)$$

Moreover, the posterior probability $\pi^{y,s}$ is continuous in y with respect to the Hellinger metric.

Remark 5.1.1. *In the original work for the definition of the above space, they use any α-Hölder within $\alpha \in (0,1]$ instead of $C^{0,1}([0,1])$ and the appropriate norm for each choice of α. Also, the statement in the original theorem includes the choice of α.*

Due to the computational complexity of the likelihood, we can easily see that the estimation of the posterior is becoming very expensive in terms of computational cost. Observe that the implementation of a Monte Carlo simulation requires the evaluation of the likelihood in several points (μ, σ) which means that for every single choice of (μ, σ), we need to compute the solution of the corresponding equation (5.6). For this reason, it is important to find an approximation that reduces the computational cost for the evaluation of the likelihood.

In the last mentioned paper, they provide a quite general theorem for the approximation of the posterior π^y. For example, that theorem can be used in combination with a Galerkin method which approximates the solution of equation (5.6). That could probably accelerate the estimation of the posterior π^y.

In our case, we consider a slightly different problem, as the diffusion term depends on a large N, this kind of diffusion term appears in many population models, where also N is supposed to be very large. Based on the last observation of our problem, we consider a different type of approximation from the one considered in Croix et al. (2020). The proposing approach, for the posterior π^y, considers an approximated posterior with a much easier to compute likelihood. More specifically, we use a perturbation technique to approximate $X^{(\mu,\sigma)}(\cdot)$ with the stochastic process $\left(X_0^{(\mu,\sigma)} + \frac{1}{\sqrt{N}} X_1^{(\mu,\sigma)}\right)(\cdot)$. Considering now a new inverse problem based on the collected data y and with forward map emerging from $\left(X_0^{(\mu,\sigma)} + \frac{1}{\sqrt{N}} X_1^{(\mu,\sigma)}\right)(\cdot)$, we end up with a new likelihood which is much easier to compute.

In the following sections, we firstly discuss the perturbation technique and provide the convergence of that technique to the solution $X^{(\mu,\sigma)}(\cdot)$. Then, we state the approximated formulation for the Bayesian inverse problem using the process $\left(X_0^{(\mu,\sigma)} + \frac{1}{\sqrt{N}} X_1^{(\mu,\sigma)}\right)(\cdot)$. The approximated posterior is denoted by ν_N^y, also note that both . Then, we prove the well-posedness of ν_N^y, and finally, we show that for large N, π_N^y and ν_N^y are close under the Hellinger metric. Also, in most of the following notes the index N on the probabilities π_N^y and ν_N^y may be omitted from time to time without prejudice.

5.2 Perturbation technique

In the case of an Itô diffusion process with fine scale diffusion coefficient, one can use a perturbation technique for the approximation of that process. We start with the presentation of that perturbation technique for Itô diffusion process with no boundary conditions. Afterwards, we suggest a similar approximation for Itô diffusion process with a reflecting barrier. Our contribution is the establishment of the approximations' convergence for an Itô diffusion process specifically with the derivative of the drift coefficient to be Lipschitz continuous and the diffusion coefficient to be 1/2-Hölder continuous.

Let us start with the simple case where $X : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is an Itô diffusion process with no boundary conditions as in Section 3.1. Suppose process X with fine scale diffusion coefficient satisfies the following stochastic differential equation

$$dX(t) = \mu\left(X(t)\right) dt + \frac{1}{\sqrt{N}} \sigma\left(X(t)\right) dW_t, \quad \forall t \geq 0 \text{ and } X(0) = x_0 \tag{5.8}$$

with parameter N to be a large number.

Our aim in this technique is to show that there are simpler processes X_0, X_1 which approximate X as N tends to infinity. For the purpose of this section an extra assumption has to be made, there exists N_0 such that for every $N \geq N_0$ the corresponding equation admits a solution. The intuition behind this approximation is that one wants to find processes $X_0, X_1 \ldots$ which are independent of N and approximate X_t in the following way

$$X_0(t) + \frac{1}{\sqrt{N}} X_1(t) + \frac{1}{N} X_2(t) + \dots, \qquad \text{for every } N \geq N_0 \tag{5.9}$$

If we further assume that coefficients μ, σ are sufficiently smooth for applying Taylor expansion on them, then by substituting the above series into equation (5.8), we obtain that

$$\begin{aligned} dX_0 + \frac{1}{\sqrt{N}}\, dX_1 + \frac{1}{N}\, dX_2 \ \dots = & \\ \mu(X_0)\, dt + \frac{1}{\sqrt{N}}\, \mu'(X_0) X_1\, dt &+ \frac{1}{N}\, \left(\mu'(X_0) X_2 + \tfrac{1}{2}\mu''(X_0) X_1^2\right)\, dt \ \dots \\ &+ \frac{1}{\sqrt{N}}\, \sigma(X_0)\, dW_t + \frac{1}{N}\, \sigma'(X_0) X_1 dW_t + \dots \end{aligned}$$

Since $X_i(t)$ does not depend on N, for any i, and the above equation holds for every $N \geq N_0$, we can obtain that X_0, X_1, $\dots$ satisfies the following equations

$$dX_0 = \mu(X_0)\, dt, \qquad dX_1 = \mu'(X_0) X_1\, dt + \sigma(X_0)\, dW_t, \dots$$

with initial conditions $X_0(0) = x_0$, $X_1(0) = 0$, $X_2(0) = 0, \dots$

In our approximation, we only consider the first two processes, i.e. X_0 and X_1, which are defined through the following differential equations

$$\begin{aligned} dX_0(t) &= \mu\left(X_0(t)\right) dt && \text{with } X_0\left(0\right) = x_0 \\ dX_1(t) &= \mu'\left(X_0(t)\right) X_1(t) dt\ + \sigma\left(X_0(t)\right) dW_t && \text{with } X_1\left(0\right) = 0 \end{aligned} \tag{5.10}$$

and our approximation can be written as it follows

$$\tilde{X}(t) = X_0(t) + \frac{1}{\sqrt{N}} X_1(t) \tag{5.11}$$

note also that $\tilde{X}(t)$ satisfies the following equation

$$\begin{aligned} &d\tilde{X}(t) = \Big(\mu(X_0(t)) + \mu'(X_0(t))(\tilde{X}(t) - X_0(t))\Big)\, dt + \frac{1}{\sqrt{N}} \sigma(X_0(t)) \mathrm{d}W_t, \\ &\tilde{X}(0) = x_0 \end{aligned} \tag{5.12}$$

Note also that the existence and uniqueness of deterministic process X_0 follows from the Lipschitz condition assumed for μ. Hence, X_0 is a well defined continuous function, and on every close interval $[0, T]$, X_0 is bounded. Thus, it is easy to see that the drift and the diffusion term of X_1 are satisfying the Lipschitz and linear growth condition. Hence, there exists a unique solution for X_1, see for instance Theorem 3.1.1. Therefore, the process $\tilde{X}$ is well-defined.

The following theorem shows that under appropriate assumptions on drift and diffusion coefficients and for large N, the process $\tilde{X}$ approaches X. More precisely, we prove that

$$X = \tilde{X} + O\left(\frac{1}{N^a}\right)$$

where $O(\cdot)$ is defined based on the norm $\mathbb{E}\sup|\cdot|^p$ and as we will show in the following theorem $a \geq 3/4$. For the sake of clarity of the following result, we define γ

$$\gamma(t,N) := N\left(X(t) - \tilde{X}(t)\right) \tag{5.13}$$

the so-called residual term of the approximation.

Theorem 5.2.1. *We consider equations (5.8) and (5.10) with coefficients μ and σ, and γ as it is defined in (5.13). Suppose μ, μ' and σ satisfy the Lipschitz condition. Then, it holds that:*

$$\mathbb{E}\sup_{\substack{t\in[0,T]\\ N\geq N_0}} |\gamma(t,N)|^p < \infty, \qquad \text{whenever } T>0 \text{ and } p\geq 1$$

Furthermore, if we only have that σ is $\frac{1}{2}$-Hölder, then it holds that:

$$\mathbb{E}\sup_{\substack{t\in[0,T]\\ N\geq N_0}} \left|\frac{1}{\sqrt[4]{N}}\gamma(t,N)\right|^p < \infty, \qquad \text{whenever } T>0 \text{ and } p\geq 1$$

The proofs of the two above results are very similar. For the proof of the first part, one can find the original proof of Theorem 3 in Blagoveshchenskii (1962). The proof of the second part is a modification of the original one and it is given in Section 5.4.2.

5.2.1 Approximated SDE with Reflecting barrier

In the current Section, we define stochastic process ξ defined on the positive half-line $[0,\infty)$ with reflecting barrier at point 0, which also satisfies equation (5.8) as long $\xi(t)$ takes values in $(0,\infty)$. As it discussed in Section 3.2.2, this kind of stochastic process ξ can always be defined as the pair of stochastic processes (l,ξ) which satisfies the following equation

$$d\xi(t) = \mu\left(\xi(t)\right)dt + \frac{1}{\sqrt{N}}\sigma\left(\xi(t)\right)dW_t + dl(t), \qquad \xi(0) = x_0 \tag{5.14}$$

for a more detailed definition of the pair (l,ξ), one can see Definition 3.2.2.

In addition, we remind that the definition of (l,ξ) is related to the Skorokhod problem, see Definition 3.2.1. Furthermore, as it discussed in Section 3.2.2, one can always define an equivalent differential form for the definition ξ. Let us consider stochastic process Y which satisfies the following SDE

$$Y(t) = \xi_0 + \int_0^t \mu(\Gamma Y(s))ds + \int_0^t \sigma(\Gamma Y(s))dW_s, \qquad Y(0) = x_0 \tag{5.15}$$

where $\Gamma : C([0,T]) \to C([0,T])$ is the Skorokhod map with $\Gamma f(t) = f(t) - \min_{s\in[0,t]}\{f(s)\wedge 0\}$ and $\xi(t) := \Gamma Y(t)$. For further properties of map Γ, one can see Lemma 3.2.1.

Inspired by the approximation demonstrated in the last Section, we are interested in defining an approximation $\tilde{\xi}$ in a similar way, as we did for $\tilde{X}$. Our first step is to define the analogue of X_0, denoted by ξ_0. For the definition of such function, we consider a non-negative function ξ_0 which satisfies equation (5.10) as long $\xi_0(t)$ takes values in $(0, \infty)$ and also has a reflective barrier at point 0. Thus, (ξ_0, l_0) is defined as the pair of solutions which satisfies equation

$$d\xi_0(t) = \mu(\xi_0(t))dt + dl_0(t), \qquad \xi_0(0) = x_0. \tag{5.16}$$

Then, according to Theorem 2 in Anderson and Orey (1976), ξ_0 satisfies the following limit

$$\mathbb{E} \sup_{t \in [0,T]} |\xi(t) - \xi_0(t)|^p \to 0, \qquad \text{as } N \to \infty, \text{ for any } p \geq 1$$

Let us now define the approximation $\tilde{\xi}$, for the stochastic process ξ, to be a non-negative process with with reflecting barrier at point 0 which satisfies equation (5.12) as long $\tilde{\xi}(t)$ takes values in $(0, \infty)$. For the definition of which, we are going to consider a pair of solutions $\left(\tilde{l}, \tilde{\xi}\right)$ which satisfies the following SDE

$$d\tilde{\xi}(t) = \Big(\mu(\xi_0(t)) + \mu'(\xi_0(t))(\tilde{\xi} - \xi_0(t)) \Big) dt + \frac{1}{\sqrt{N}} \sigma(\xi_0(t)) dW_t + d\tilde{l}(t), \qquad \tilde{\xi}(0) = x_0 \tag{5.17}$$

For completeness, we may then define the analogue of X_1, denoted by ξ_1, but in this case, ξ_1 is not necessarily a non-negative process with with reflecting barrier at point 0. The definition of which comes from the following expression $\xi_1 := N\left(\tilde{\xi} - \xi_0\right)$, and we can easily see that ξ_1 satisfies the following equation,

$$d\xi_1(t) = \mu'(\xi_0(t))\xi_1(t)dt + \sigma(\xi_0(t))dW_t + \sqrt{N} d(\tilde{l} - l_0)(t), \qquad \xi_1(0) = 0 \tag{5.18}$$

Similarly as we did for the case of the unbounded stochastic process $\tilde{X}$, we can show that $\tilde{\xi}$ is well-defined, but first, we need to define the following two equivalent differential forms of ξ_0 and $\tilde{\xi}$. We first consider processes Y_0 and $\tilde{Y}$ which satisfy the following two equations

$$\begin{aligned} dY_0(t) &= \mu(\Gamma(Y_0)(t))dt \\ d\tilde{Y}(t) &= \mu(\Gamma(Y_0)(t)) + \mu'(\Gamma(Y_0)(t)) \Big(\Gamma(\tilde{Y})(t) - \Gamma(Y_0)(t) \Big) dt + \frac{1}{\sqrt{N}} \sigma\left(\Gamma(Y_0)(t)\right) dW_t \end{aligned} \tag{5.19}$$

with initial values $Y_0(0) = x_0$ and $\tilde{Y}(0) = x_0$, and then, we can define ξ_0 and $\tilde{\xi}$ as follows: $\xi_0(t) = \Gamma(Y_0)(t)$ and $\tilde{\xi}(t) = \Gamma(\tilde{Y})(t)$.

Note then that Y_0 is well-defined, since $\mu(\Gamma(\cdot)(t))$ satisfies the Lipschitz condition and the linear growth. Then, using the existence and uniqueness, see Theorem 3.2.2, for the

Skorokhod problem, we get that the pair (l_0, ξ_0) is well-defined. Last, using the boundedness of ξ_0 on a bounded interval $[0, T]$, we see that the drift and diffusion coefficients of $\tilde{\xi}$ satisfy the hypothesis assumed in Theorem 3.2.3. Thus, we have that the pair $\left(\tilde{l}, \tilde{\xi}\right)$ is well-defined.

The following Theorem is analogous to Theorem 5.2.1 for the case where ξ is a fine scale diffusion coefficient with reflecting barrier at zero. The proof of the Theorem can found in Section 5.4.2. Let us also re-consider and re-write the definition of the residual term for this case,

$$\gamma(t, N) := N\left(\xi(t) - \tilde{\xi}(t)\right) \tag{5.20}$$

Theorem 5.2.2. *We consider equations* (5.14) *and* (5.17) *with coefficients* μ *and* σ*, and* γ *as it is defined in* (5.20)*. Suppose* μ*,* μ' *satisfy the Lipschitz condition and* σ *is* $\frac{1}{2}$*-Hölder continuous. Then, it holds that:*

$$\mathbb{E} \sup_{\substack{t\in[0,T]\\ N\geq N_0}} \left| \frac{1}{\sqrt[4]{N}} \gamma(t, N) \right|^p < \infty, \qquad \textit{whenever } T > 0 \textit{ and } p \geq 1$$

5.3 Likelihood of the approximate posterior

Using the approximate stochastic process of the previous section in our Bayesian inverse problem, for finding the drift and diffusion terms, gives rise to an approximation of the posterior which is the subject of this section. At this point, we need to mention that the results of the current Section concern only the case where the stochastic process X has no boundary conditions, i.e. X has neither absorbing nor reflecting barrier on the boundaries. Therefore, the posterior distribution π^y and its approximation ν^y correspond to the Bayesian inverse problem with forward map the unrestricted stochastic process X. The proof for the case of the stochastic process with absorbing and reflective barrier is still in progress.

Let us consider an appropriate function space for the unbounded stochastic process: given constants $C_1, C_2, C_3 > 0$, then we define

$$\Lambda_{C_1,C_2,C_3} = \left\{(\mu, \sigma) : \mu \in C^{1,1}(\mathbb{R}), \sigma^2 \in C^{0,1}(\mathbb{R}), \min_{x\in\mathbb{R}} \sigma(x) \geq C_1, \|\sigma^2\|_{0,1} \leq C_2, \|\mu\|_{1,1} \leq C_3\right\}$$

we also recall the definition of $\|\cdot\|_{1,1}$ which is given as follows: given $f \in C^{1,1}(\mathbb{R})$

$$\|f\|_{1,1} := \|f\|_{L^\infty(\mathbb{R})} + \|f'\|_{L^\infty(\mathbb{R})} + \sup_{x,y\in\mathbb{R}} \frac{\|f'(x) - f'(y)\|}{\|x - y\|}.$$

For the approximation of the posterior π^y, we suggest the probability measure ν^y over the function space Λ_{C_1,C_2,C_3}. We first define process $\tilde{X}^{(\mu,\sigma)}(t; \xi, \tau)$ based on the framework built in the previous section. In addition, we denote by $q^{(\mu,\sigma)}_{\xi,\tau}$ the transition density function of $\tilde{X}^{(\mu,\sigma)}(t;\xi,\tau)$, and then, we define function $Q(y,t;\xi,\tau) := q_{\xi,\tau}(y,t;\xi,\tau)$. Using the definition of Q, we then can define ν^y through its Radon-Nikodym with respect to π_0, similarly as in the case of π^y, see for instance (5.5),

$$\frac{d\nu^y}{d\pi_0}(\mu,\sigma) \propto \tilde{\mathcal{L}}^y(\mu,\sigma) = \prod_{i=0}^{J-1} \tilde{Q}^{(\mu,\sigma)}(\tilde{y}_{i+1},t_{i+1};\tilde{y}_i,t_i), \tag{5.21}$$

where $\tilde{Q}^{(\mu,\sigma)}(y,t;x,s)$ is defined as follows

$$\tilde{Q}^{(\mu,\sigma)}(y,t;x,s) := \int\int Q^{(\mu,\sigma)}(z,t;w,s) f_{N(0,\epsilon^2)}(y-z)\, f_{N(0,\epsilon^2)}(x-w) dz dw. \tag{5.22}$$

Let us then take $\mu \in C^{1,1}(\mathbb{R})$ and $\sigma^2 \in C^{0,1}(\mathbb{R})$, and consider $\tilde{X}$ similarly as in equation (5.12) with the only difference that we know the initial values of $\tilde{X}$, X_0 and X_1 at time τ instead of time 0. Since, $\tilde{X}$ depends on μ and σ, it is more convenient to change the notation of X_0 to $\phi_\tau(t;\mu,\xi)$, or simply $\phi_\tau(t;\xi)$, where parameters τ, ξ correspond to the initial value of ϕ, i.e. $\phi_\tau(\tau;\xi) = \xi$. Similarly for X_1, we may use the parameter τ, i.e. $X_1(t;\tau)$, to highlight that X_1 is the process with initial value 0 at time τ, i.e. $X_1(\tau;\tau) = 0$. Let us now rewrite equations (5.11) and (5.12) in terms of $\phi_\tau(t;\mu,\xi)$ and $X_1(t;\tau)$

$$\tilde{X}(t;\xi,\tau) = \phi_\tau(t;\mu,\xi) + \frac{1}{\sqrt{N}} X_1(t;\tau) \tag{5.23}$$

where $\phi_\tau(t;\mu,\xi)$ and $X_1(t;\tau)$ satisfy equations (5.10) with initial value at τ. Next, we have that $\tilde{X}(t;\xi,\tau)$, or simply $\tilde{X}(t)$, satisfies the following equation

$$\begin{aligned} d\tilde{X}(t) &= \tilde{\mu}(\tilde{X}(t),t)dt + \frac{1}{\sqrt{N}}\tilde{\sigma}(t)\mathrm{d}W_t, \qquad \text{for } t \geq t_0 \\ \tilde{X}(\tau) &= \xi \end{aligned} \tag{5.24}$$

where $\tilde{\mu}$ and $\tilde{\sigma}$ are defined as follows

$$\begin{aligned} \tilde{\mu}(x,t) &= \mu\left(\phi_\tau(t;\xi)\right) + \mu'\left(\phi_\tau(t;\xi)\right)\left(x - \phi_\tau(t;\xi)\right) \\ \tilde{\sigma}(t) &= \sigma\left(\phi_\tau(t;\xi)\right). \end{aligned}$$

Let us now obtain the transition probability of $\tilde{X}$. We see that equation (5.10) can be also be written as follows

$$\mathrm{d}\left(X_1(t)\,\mathrm{e}^{-\int_{t_0}^t \mu'(\phi_\tau(\lambda;\xi))\mathrm{d}\lambda}\right) = \sigma\left(\phi_\tau(t;\xi)\right)\,\mathrm{e}^{-\int_{t_0}^t \mu'(\phi_\tau(\lambda;\xi))\mathrm{d}\lambda}\,\mathrm{d}W_t$$

Or equivalently, we have that for $t_0 \leq s \leq t$

$$X_1(t) = X_1(s)\mathrm{e}^{\int_s^t \mu'(\phi_\tau(\lambda;\xi))\mathrm{d}\lambda} + \int_s^t \sigma\left(\phi_\tau(q;\xi)\right)\,\mathrm{e}^{\int_q^t \mu'(\phi_\tau(\lambda;\xi))\mathrm{d}\lambda}\,\mathrm{d}W_q \tag{5.25}$$

in Section 5.4.1, one can find more details on how we obtain the last equation.

Next, if we apply the explicit form of X_1 into the expression of (5.23), we get that

$$\tilde{X}(t) = \phi_\tau\,(t;\xi) + \Big(\tilde{X}(s) - \phi_\tau\,(s;\xi)\Big)\,\mathrm{e}^{\int_s^t \mu'(\,\phi_\tau(\lambda;\xi)\,)\mathrm{d}\lambda} + \frac{1}{\sqrt{N}}\int_s^t \sigma\,(\,\phi_\tau(s;\xi)\,)\,\mathrm{e}^{\int_q^t \mu'(\,\phi_\tau(\lambda;\xi)\,)\mathrm{d}\lambda}\,\mathrm{d}W_q \tag{5.26}$$

Hence, we have that the transition probability of $\tilde{X}^{(\mu,\sigma)}(t;\xi,\tau)$ is Gaussian distributed, i.e.

$$q_{\xi,\tau}^{(\mu,\sigma)}(y,t;x,s) := \frac{1}{\sqrt{2\,\pi F_N(t;s)}}\mathrm{e}^{-\frac{(y-\Phi(t;x,s))^2}{2F_N(t;s)}} \tag{5.27}$$

where $\Phi\,(t;x,s)$ and $F_N(t;s)$ are defined as follows

$$\Phi\,(t;x,s) = \phi_\tau\,(t;\xi) + (x - \phi_\tau\,(s;\xi))\,\mathrm{e}^{\int_s^t \mu'(\,\phi_\tau(\lambda;\xi)\,)\mathrm{d}\lambda}$$
$$F_N(t;s) = \frac{1}{N}\int_s^t \sigma^2\,(\,\phi_\tau(q;\xi)\,)\;\mathrm{e}^{2\int_q^t \mu'(\,\phi_\tau(\lambda;\xi)\,)d\lambda}\,dq,$$

observe that $\Phi\,(t;x,s)$ and $F_N(t;s)$ depend on ξ and τ, but for the readability of the current work, we choose not to include them in our notation. We also note that $q_{\xi,\tau}(y,t;x,s)$ satisfies the associated backward Kolmogorov equation of $\tilde{X}(t;\xi,\tau)$,

$$\begin{aligned} -\frac{dq_{\xi,\tau}\,(x,s)}{ds} &= \frac{\tilde{\sigma}^2(s)}{2N}\frac{\partial^2 q_{\xi,\tau}\,(x,s)}{\partial x^2} + \tilde{\mu}(x,s)\frac{\partial q_{\xi,\tau}\,(x,s)}{\partial x}, \\ \lim_{s\nearrow t} q_{\xi,\tau}&(y,t;x,s) = \delta(y-x), \end{aligned} \tag{5.28}$$

variables (y,t) are omitted in the above expression, since these are fixed.

Since we have obtained the unnormalised likelihood $\tilde{\mathcal{L}}^y\,(\mu,\sigma)$, we can now state the following Theorem which provides us that the proposed measure ν^y is well-defined. Its proof can be found in Section 5.4.2.

Theorem 5.3.1. *Let T>0, $J \geq 1$ and consider $y = (y_1,\dots,y_J) \in \mathbb{R}^J$ to be the data at time $t = (t_1,\dots,t_J)$ with $0 < t_1 < \cdots < t_J \leq T$. Suppose a probability measure π_0 and set $C_M := \left\{(\mu,\sigma) \in \Lambda_{C_1,C_2,C_3} : \max_{i\in\{0,\dots J-1\}} |y_{i+1} - \phi_{t_i}\,(t_{i+1};\mu,y_i)| \leq M\right\}$ such that $\pi_0(\Lambda_{C_1,C_2,C_3}) = 1$ and $\pi_0\,(C_M) > 0$. Then there exists a unique measure ν^y given by*

$$\frac{d\nu^y}{d\pi_0}(\mu,\sigma) = \frac{1}{\tilde{Z}(y)}\tilde{\mathcal{L}}^y\,(\mu,\sigma)$$

where $\tilde{Z}(y)$ is the normalisation constant given as

$$\tilde{Z}(y) = \int_\Lambda \tilde{\mathcal{L}}^y\,(\mu,\sigma)\,\mathrm{d}\pi_0\,(\mu,\sigma) \in (0,\infty)$$

Remark 5.3.1. *The last theorem can be applied on priors π_0 which have bounded support similar to the example of uniform priors as in Section 2.3.1. Similar to the uniform priors, one can define r-exponential prior distributions, $r \geq 1$, on bounded subsets of Λ_{C_1,C_2,C_3}, by using for example truncated r-exponential distributions on ξ, see the model (2.7) in Section 2.3.*

5.3.1 Convergence of the approximated posterior

In this part, we provide an estimate for the Hellinger distance between the approximated posterior ν^y and the posterior π^y as N tends to infinity. The aim of which is to finally prove the convergence of ν^y to the posterior π^y as N tends infinity.

Essential component for estimating the above mentioned distance is the following theorem. The proof of the following theorem can be found in Section 5.4.2

Theorem 5.3.2. *Given $(\mu, \sigma) \in \Lambda_{C_1,C_2,C_3}$, we consider the transition density function p of Itô diffusion X with X satisfies equation (5.8), and the transition density function Q of Itô diffusion $\tilde{X}(t;\xi,\tau)$ with $\tilde{X}(t;\xi,\tau)$ as it is defined in equation (5.24). Then, for every $\xi \in \mathbb{R}$, it holds that*

$$\left|p *_t f_{N(0,\epsilon^2)}(y,t;\xi,\tau) - Q *_t f_{N(0,\epsilon^2)}(y,t;\xi,\tau)\right| \leq \frac{C\left(1+\xi^2\right)}{\sqrt{N}} \tag{5.29}$$

with C depends on $t-\tau$, ϵ, C_1, C_2 and C_3.

Remark 5.3.2. *Using the well-posedness of the transformation which is used for the reflective barrier (Skorohod map), see Section 3.2.2, and an appropriate stopping time for the left hand side Dirichlet boundary condition, one can show this result for the bounded interval.*

We now fix the prior to obtain the approximation rates. We assume that π_0 is the truncated α-regular r-exponential measure, with $r \geq 1$ and $\alpha \geq 4$, that is, the draws u of π_0 are defined as

$$u = \sum_{j=1}^{\infty} j^{-\frac{1}{2}-\alpha}\xi_j\psi_j \tag{5.30}$$

with $\{\psi_j\}$ a regular enough basis for Λ_{C_1,C_2,C_3} and ξ_j appropriately truncated i.i.d r-exponential random variable with density $c_r \mathrm{e}^{-|x|^r/r}$ (see Agapiou et al. (2021)). The following theorem provides an estimate for the Hellinger distance between the approximated posterior ν^y and the posterior π^y. Its proof can be found in Section 5.4.2.

Theorem 5.3.3. *Let us consider the same set-up as in Theorem 5.3.1. Then, it holds that*

$$d_{Hell}(\pi^y, \nu^y) \leq \frac{C}{\sqrt{N}}$$

for large N and with C depend on C_1, C_2, C_3, J, M, ϵ, $\pi_0\left(C_M\right)$ and $\Delta t := \max\limits_{i\in\{1,\ldots J\}} |t_i - t_{i-1}|$.

5.4 Proofs for Chapter 5

5.4.1 Explicit form of the approximated solution

For given X_0, the second equation in (5.10) can be written as follows:

$$dX_1 = g(t)X_1dt + b(t)dW_t$$

where g and b are continuous functions. Let us define $G(t) = e^{-\int_{t_0}^{t} g(s)ds}$, the derivative of which is given as follows: $G'(t) = -g(t)G(t)$. Let us take integral of G with respect to dX_1,

$$G(t)dX_1 = g(t)G(t)X_1dt + b(t)G(t)dW_t$$

Equivalently, we have that

$$G'(t)X_1\, dt + G(t)dX_1 = b(t)G(t)dW_t$$

Applying Itô's formula on the product $G(t)X_t$, we get that:

$$d\left(G(t)X_1\right) = G'(t)X_1\, dt + G(t)dX_1$$

Using the last two equations, we get that:

$$G(t)X_1(t) - G(t_0)X_1(t_0) = \int_{t_0}^{t} G(s)b(s)dW_s$$

The initial value of $X_1(t_0) = 0$, thus we obtain the following solution:

$$X_1(t) = \frac{1}{G(t)}\int_0^t G(s)b(s)dW_s$$

Therefore, we get:

$$X_1(t) = \int_0^t b(s)\, e^{\int_s^t g(q)dq}\, dW_s$$

5.4.2 Proofs of the theorems

Proof of theorem 5.2.1

We first show the desired result for $p \geq 2$. Let us first denote the following function $M(\tilde{X}(t), t) := \mu(X_0(t)) + \mu'(X_0(t))(\tilde{X}(t) - X_0(t))$ and consider γ according to (5.13), for the definition of X and $\tilde{X}$, see equations (5.8) and (5.11),

$$\frac{1}{N}\gamma(t, N) = \int_0^t \mu\left(X(q)\right) - M(\tilde{X}(q), q)\; dq + \int_0^t \sigma\left(X(q)\right) - \sigma\left(X_0(q)\right)\; dW_q$$

Next, we only consider the term which appears in the first integral

$$\left|\mu\left(X(q)\right)-M(\tilde{X}(q),q)\right| \leq \left|\mu\left(X(q)\right)-\mu\left(\tilde{X}(q)\right)\right| + \left|\mu\left(\tilde{X}(q)\right)-M(\tilde{X}(t),t)\right| \tag{5.31}$$

Using the mean value Theorem, we can obtain $\theta(q) \in [0, \frac{1}{\sqrt{N}}]$ for every q such that:

$$\mu\Big(X_0(q)+\tfrac{1}{\sqrt{N}}X_1(q)\Big)-\mu\left(X_0(q)\right)=\tfrac{1}{\sqrt{N}}X_1(q)\,\mu'\left(X_0(q)+\theta(q)X_1(q)\right)$$

We observe that by the definition of $\tilde{X}$, the difference $\tilde{X}(t)-X_0(t)=\frac{1}{\sqrt{N}}X_1(t)$. In addition, we note that the above difference appears in the second term of the right hand side of inequality (5.31). Thus, if we apply the Lipschitz property of μ and μ', we get that

$$\begin{aligned}\left|\mu\left(X(q)\right)-M(\tilde{X}(q),q)\right| &\leq \|\mu\|_{0,1}\left|X(q)-\tilde{X}(q)\right| \\ &\qquad + \frac{1}{\sqrt{N}}\left|X_1(q)\right|\left|\,\mu'\left(X_0(q)+\theta(q)X_1(q)\right)-\mu'\left(X_0(q)\right)\right| \\ &\leq \frac{\|\mu\|_{0,1}}{N}\left|\gamma(q,N)\right|+\frac{\|\mu'\|_{0,1}}{N}\left|X_1(q)\right|^2\end{aligned} \tag{5.32}$$

we remind that $\gamma(q,N) := N\left(X(q)-\tilde{X}(q)\right)$. Hence, γ can be bounded from the following

$$\begin{aligned}\frac{1}{N}\gamma(t,N) \leq \frac{\|\mu\|_{1,1}}{N}&\left[\int_0^t \sup_{s\in[0,q]}|\gamma(s,N)| + \sup_{s\in[0,q]} X_1^2(s)dq\right] \\ &+\frac{1}{\sqrt{N}}\int_0^t \sigma\left(X(q)\right)-\sigma\left(X_0(q)\right)dW_q\end{aligned} \tag{5.33}$$

See now that $|x|^p$ is convex, thus Jensen inequality implies that: $(a+b)^p \leq 2^{p-1}\left(a^p+b^p\right)$, whenever $a,b\geq 0$. Next, we consider the expectation of $\gamma(t,N)$, the following inequality is obtained by applying both inequality (5.33) and the above property of $|x|^p$

$$\begin{aligned}\mathbb{E}\sup_{0\leq s\leq t}|\tfrac{1}{N}\gamma(s,N)|^p \lesssim \frac{t^{p-1}}{N}\mathbb{E}\int_0^t |\gamma(q,N)|^p + \sup_{s\in[0,q]} X_1^{2p}(s)\,dq \\ +\frac{1}{N^{p/2}}\mathbb{E}\sup_{0\leq s\leq t}\left|\int_0^t \sigma\left(X(q)\right)-\sigma\left(X_0(q)\right)dW_q\right|^p\end{aligned}$$

observe also that the last inequality and more specifically, in order to obtain the first term of the right hand side of the above inequality, we need to apply the Hölder inequality.

According to Lemma 5.4.1, we have that the expected value of the supremum of X_1^{2p} is bounded. Moreover, we have that the expected value of the supremum of X for every $p\geq 2$ is bounded. See that both coefficients $\mu(\cdot)$ and $\sigma(\cdot)$ are satisfying the linear growth condition, since $\mu(\cdot)$ is Lipschitz and $\sigma(\cdot)$ is $\frac{1}{2}$-Hölder continuous and depends only on x and not on t. Then, the boundedness of X follows from Lemma 2.3.2 in Mao (2007). The upper bound of the integral with respect to the Brownian motion comes from Theorem 1.7.2 in Mao (2007)

$$\mathbb{E}\sup_{0\le s\le t}|\tfrac{1}{N}\gamma(s,N)|^p \lesssim \frac{1}{N^p}\left[\int_0^t \mathbb{E}\sup_{0\le s\le q}|\gamma(s,N)|^p dq+1\right]+\frac{1}{N^{p/2}}\int_0^t \mathbb{E}\,|\sigma\left(X(q)\right)-\sigma\left(X_0(q)\right)|^p dq$$

Let us now consider only the last term of the right hand side of the above inequality and apply $\frac{1}{2}$-Hölder continuity of σ, we have that

$$\int_0^t \mathbb{E}\,|\sigma\left(X(q)\right)-\sigma\left(X_0(q)\right)|^p\,dq \le \left(\frac{\|\sigma\|^2_{0,\frac{1}{2}}}{\sqrt{N}}\right)^{\frac{p}{2}}\int_0^t \mathbb{E}\sup_{s\in[0,q]}\left|X_1(s)+\frac{1}{\sqrt{N}}\gamma(s,N)\right|^{\frac{p}{2}}dq \tag{5.34}$$

Next, we apply once again inequality $(a+b)^p \le 2^{p-1}(a^p+b^p)$, the boundedness of the expected value of the supremum of X_1, and also, that $1+|x|^{\frac{p}{2}} \lesssim 1+|x|^p$. Applying all three of them in the above (5.34), we get that

$$\int_0^t \mathbb{E}\,|\sigma\left(X(q)\right)-\sigma\left(X_0(q)\right)|^p dq \lesssim \frac{1}{N^{p/4}}\left[1+\int_0^t \mathbb{E}\sup_{s\in[0,q]}\left|\frac{1}{\sqrt{N}}\gamma(s,N)\right|^p dq\right]$$

Thus,

$$\frac{1}{N^p}\mathbb{E}\sup_{0\le s\le t}|\gamma(s,N)|^p \lesssim \frac{1}{N^p}\int_0^t \mathbb{E}\sup_{0\le s\le q}|\gamma(s,N)|^p\,dq+\frac{1}{N^{\frac{3p}{4}}}$$

Or, equivalently,

$$\mathbb{E}\sup_{0\le s\le t}\left|\frac{1}{\sqrt[4]{N}}\gamma(s,N)\right|^p \lesssim \int_0^t \mathbb{E}\sup_{0\le s\le q}\left|\frac{1}{\sqrt[4]{N}}\gamma(s,N)\right|^p dq+1$$

Using Grönwall's inequality, see for instance Lemma1.1[1] on page 30 in Freidlin and Wentzell (2012), implies that,

$$\mathbb{E}\sup_{0\le s\le t}|\frac{1}{\sqrt[4]{N}}\gamma(s,N)|^p \le 2Ce^{Ct}$$

note also that C does not depend on N.

For $1\le p<2$ the result follows from an application of Hölder inequality.

Proof of Theorem 5.2.2

The proof of this Theorem is quite similar to the proof of Theorem 5.2.1. We first show the desired result for $p\ge 2$. As we discuss in Section 5.2.1, we can define Y_0 and $\tilde{Y}$ which satisfy equations (5.19), then ξ_0 and $\tilde{\xi}$ can be defined through Skorokhod map as it follows $\xi_0(t)=\Gamma(Y_0)(t)$ and $\tilde{\xi}(t)=\Gamma(\tilde{Y})(t)$. Let us denote the following function $M(\tilde{\xi}(t),t):=\mu(\xi_0(t))+\mu'(\xi_0(t))(\tilde{\xi}(t)-\xi_0(t))$ and consider γ according to (5.20) and Y as

[1]If $m(t)\le a\int_0^t m(s)ds+C$ implies $m(t)\le Ce^{at}$

in equation (5.15) with $\xi(t) = \Gamma(Y)(t)$. Then, let us take the supremum of γ,

$$\begin{aligned}
\sup_{s\in[0,t]} \left|\frac{1}{N}\gamma(s,N)\right| &= \sup_{s\in[0,t]} \left|\xi(t) - \tilde{\xi}(t)\right| \leq 2 \sup_{s\in[0,t]} \left|Y(t) - \tilde{Y}(t)\right| \\
&= 2 \sup_{s\in[0,t]} \left|\int_0^t d\left(Y(q) - \tilde{Y}(q)\right)\right| \\
&\lesssim \sup_{s\in[0,t]} \left|\int_0^s (\mu(\xi(q)) - M(\tilde{\xi}(q), q)\,)dq\right| + \frac{1}{\sqrt{N}} \sup_{s\in[0,t]} \left|\int_0^s \sigma(\xi(q)) - \sigma(\xi_0(q))dW_q\right| \\
&\lesssim \int_0^t \left|\mu(\xi(q)) - M(\tilde{\xi}(q), q)\,\right| dq + \frac{1}{\sqrt{N}} \sup_{s\in[0,t]} \left|\int_0^s \sigma(\xi(q)) - \sigma(\xi_0(q))dW_q\right|
\end{aligned}$$

note that the first inequality is a property of the Skorokhod map, see part (i) in Lemma 3.2.1. We recall the following inequality which is immediate consequence of Jensen's inequality: $(a+b)^p \leq 2^{p-1}\left(a^p + b^p\right)$, whenever $a, b \geq 0$. Then, we take the square of the last supremum and apply Young's inequality and Hölder inequality on the first integral

$$\begin{aligned}
\sup_{s\in[0,t]} \left|\frac{1}{N}\gamma(s,N)\right|^p \lesssim t^{p-1} \int_0^t \left|\mu(\xi(q)) - M(\tilde{\xi}(q), q)\,\right|^p dq \\
+ \frac{1}{N^{p/2}} \sup_{s\in[0,t]} \left|\int_0^s \sigma(\xi(q)) - \sigma(\xi_0(q))dW_q\right|^p
\end{aligned}$$

After that, we consider the expectation of the above supremum and we combine it with Theorem 1.7.2 from Mao (2007), we can get an easier form to estimate

$$\begin{aligned}
\mathbb{E} \sup_{s\in[0,t]} \left(\frac{1}{N}\gamma(s,N)\right)^p \lesssim \mathbb{E} \int_0^t \left|\mu(\xi(q)) - M(\tilde{\xi}(q), q)\,\right|^p dq \\
+ \frac{1}{N^{p/2}}\, \mathbb{E} \int_0^t \left|\sigma(\xi(q)) - \sigma(\xi_0(q))\right|^p dq
\end{aligned} \tag{5.35}$$

we also note that the coefficient $\sqrt{t}$ has been absorbed by "$\lesssim$". Part of the following steps are similar to the proof of Theorem 5.2.1, thus some explanations are omitted. Similarly to inequality (5.32), we get that

$$\left|\mu\left(\,\xi(s)\,\right) - M(\tilde{\xi}(s), s)\right| \leq \frac{\|\mu\|_{1,1}}{N}\left(|\gamma(s,N)| + \xi_1^2(s)\right)$$

By applying inequality $(a+b)^p \leq 2^{p-1}\left(a^p + b^p\right)$, whenever $a, b \geq 0$, we get that

$$\left|\mu\left(\,\xi(s)\,\right) - M(\tilde{\xi}(s), s)\right|^p \lesssim \frac{\|\mu\|_{1,1}^p}{N^p}\left(|\gamma(s,N)|^p + \xi_1^{2p}(s)\right)$$

Let us recall Young's inequality, for given a, b non-negative and $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$, we have that $ab \leq \frac{a^p}{p} + \frac{a^q}{q}$. Using both $\frac{1}{2}$-Hölder continuity of σ and Young's inequality, we get that

$$\begin{aligned}
|\sigma(\xi(s)) - \sigma(\xi_0(s))|^2 &\leq \frac{\|\sigma\|_{0,\frac{1}{2}}^2}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}|\gamma(s,N)| + |\xi_1(s)|\right) \\
&\lesssim \frac{\|\sigma\|_{0,\frac{1}{2}}^2}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\gamma^2(s,N) + |\xi_1(s)| + 1\right)
\end{aligned}$$

By applying inequality $(a+b+c)^p \leq 3^{p-1}(a^p+b^p+c^p)$, whenever $a,b,c \geq 0$, we get that

$$|\sigma(\xi(s)) - \sigma(\xi_0(s))|^p \lesssim \frac{\|\sigma\|^p_{0,\frac{1}{2}}}{N^{p/4}} \left(\frac{1}{N^{p/4}} \gamma^p(s,N) + |\xi_1(s)|^{p/2} + 1 \right)$$

By applying the last inequalities in (5.35), we get that

$$\mathbb{E} \sup_{s\in[0,t]} \left(\frac{1}{N} \gamma(s,N) \right)^p \lesssim \int_0^t \mathbb{E} \sup_{s\in[0,q]} \left| \frac{1}{N} \gamma(s,N) \right|^p dq \\ + \frac{1}{N^{3p/4}} \int_0^t \left(\mathbb{E} \sup_{s\in[0,q]} \xi_1^{2p}(s) + \mathbb{E} \sup_{s\in[0,q]} |\xi_1(s)|^{p/2} + 1 \right) dq$$

Observe that we can interchange the order of expectation and the integral, since we have non-negative functions, and then, we can apply Tonelli's Theorem. Note also that the multiplication constant of the last inequality, i.e. $\lesssim$, depends on $\|\mu\|^p_{1,1} \vee \|\sigma\|^p_{0,\frac{1}{2}}$.

According to Lemma 5.4.2, we get that the expected value of the supremum of $|\xi_1|^p$, for every power $p \geq 1$, is bounded thus we get that the integral of the last parethensis in the right hand side of the above inequality is bounded and of order $\frac{1}{N^{3/2}}$. Hence, we get that

$$\mathbb{E} \sup_{s\in[0,t]} \left(\frac{1}{\sqrt[4]{N}} \gamma(s,N) \right)^p \lesssim \int_0^t \mathbb{E} \sup_{s\in[0,q]} \left(\frac{1}{\sqrt[4]{N}} \gamma(s,N) \right)^p dq + C$$

where C depends on the bounds of ξ_1, T, p. Then the desired result is consequence of Grönwall's inequality

$$\mathbb{E} \sup_{s\in[0,t]} \left| \frac{1}{\sqrt[4]{N}} \gamma(s,N) \right|^p \leq C_1 e^{C_2 t}$$

note also that C_1, C_2 do not depend on N.

For $1 \leq p < 2$ the result follows from an application of Hölder inequality.

Proof of Theorem 5.3.1

According to Bayes' Theorem, see for instance Theorem 2.2.1, for proving that the Bayesian approach is well-defined, it is sufficient to show that $\tilde{\mathcal{L}}^y(\mu,\sigma)$ is $\mathcal{B}(\Lambda) \otimes \mathcal{B}(Y)$-measurable and $Z(y) \in (0,+\infty)$ for a.e. y.

In order to get the continuity of $Q^{(\mu,\sigma)}(z,t_{i+1};w,t_i)$ with respect to (z,w,μ,σ), we only need to see that $Q^{(\mu,\sigma)}(z,t_{i+1};w,t_i)$ is an exponential functions of z and the forward solution of the ODE, $\phi(t_{i+1};\mu,w)$, and polynomially dependent on σ; and, one can readily verify that $\phi(t_{i+1};\mu,w)$ are continuous with respect to μ and w. From equation (5.22), we have that $\tilde{Q}^{(\mu,\sigma)}(z,t_{i+1};w,t_i) = \left(Q^{(\mu,\sigma)} *_{t_{i+1}} f_{N(0,\epsilon)}\right) *_{t_i} f_{N(0,\epsilon)}(z,t_{i+1};w,t_i)$. Since, $\tilde{Q}^{(\mu,\sigma)}$ is convolutions of $Q^{(\mu,\sigma)}$ with Gaussian densities, then we have that also $\tilde{Q}^{(\mu,\sigma)}(z,t_{i+1};w,t_i)$ is continuous. From equation (5.21), we have that the likelihood is given as the product of continuous functions. It remains to show that $\tilde{Z}^y \in (0,+\infty)$.

Next, we use that $Q^{(\mu,\sigma)}(y,t;\xi,\tau)$ follows the law of $N\left(\phi_\tau\left(t;\xi\right),F_N(t;\tau)\right)$, see for instance (5.27). Thus, we get that

$$\begin{aligned} Q^{(\mu,\sigma)} *_t f_{N(0,\epsilon^2)}(y,t;\xi,\tau) &= \int f_{N(\phi_\tau(t;\xi),\,F_N(t;\tau))}(y-z)f_{N(0,\epsilon^2)}(z)dz \\ &= f_{N(\phi_\tau(t;\xi),\,F_N(t;\tau)+\epsilon^2)}(y) \end{aligned} \tag{5.36}$$

the last equality holds, due to the following property of Gaussian distributions: If we assume that $X \sim N(\mu_X,\sigma_X^2)$ and $Y \sim N(\mu_Y,\sigma_Y^2)$ with X independent of Y, then it holds

$$X + Y \sim N(\mu_X+\mu_Y,\sigma_X^2+\sigma_Y^2). \tag{5.37}$$

The proof of the above property is considered trivial, hence its proof is omitted.

From (5.36), we get that $\tilde{Q}^{(\mu,\sigma)}(y,t;\xi,\tau)$ is bounded

$$\begin{aligned} \tilde{Q}^{(\mu,\sigma)}(y,t;\xi,\tau) &= \int q^{(\mu,\sigma)} *_t f_{N(0,\epsilon^2)}(y,t;w,\tau)f_{N(0,\epsilon^2)}(\xi-w)dw \\ &\leq \frac{1}{\epsilon}\int f_{N(0,\epsilon^2)}(\xi-w)dw = \frac{1}{\epsilon} \end{aligned} \tag{5.38}$$

Then, we have that

$$\tilde{Z}^y := \int \tilde{\mathcal{L}}^y\left(\mu,\sigma\right)\mathrm{d}\pi_0\left(\mu,\sigma\right) = \int\prod_{i=1}^J \tilde{Q}^{(\mu,\sigma)}(y_{i+1},t_{i+1};y_i,t_i)\mathrm{d}\pi_0\left(\mu,\sigma\right) \leq \frac{1}{\epsilon^J}$$

Let us denote $\Delta t := \max_{i\in\{1,\ldots J\}}|t_i-t_{i-1}|$, and get $N \geq \Delta t\left(\frac{C_{\|\sigma\|_\infty}\mathrm{e}^{2C_{\|\mu'\|_\infty}\Delta t}}{\epsilon}\right)^2$. Then, we have that for every $i\in\{0,\ldots J-1\}$, it holds that

$$\epsilon^2 \leq F_N(t_{i+1};t_i) = \frac{1}{N}\int_{t_i}^{t_{i+1}} \sigma^2\left(\,\phi_{t_i}(q;y_i)\,\right)\,\mathrm{e}^{2\int_q^{t_{i+1}}\mu'\left(\,\phi_{t_i}(\lambda;y_i)\,\right)d\lambda}\,dq\ +\epsilon^2 \leq 2\epsilon^2$$

If we have that $|x| \leq \tilde{M}$ and $\sigma\in[a,b]$, then we can easily show that $\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2\sigma^2}} \geq \frac{1}{b\sqrt{2\pi}}\mathrm{e}^{-\frac{\tilde{M}^2}{2a^2}}$. Similarly, we have that for every $(\mu,\sigma)\in C_M$ it holds that

$$\frac{1}{2\epsilon\sqrt{\pi}}\mathrm{e}^{-\frac{M^2}{2\epsilon^2}} \leq f_{N\left(\phi_{t_i}(t_{i+1};y_i),\,F_N(t_{i+1};t_i)+\epsilon^2\right)}(y_{i+1}) = \tilde{Q}^{(\mu,\sigma)}(y_{i+1},t_{i+1};y_i,t_i)$$

for every $i\in\{0,\ldots J-1\}$. Therefore, we obtain that

$$0 < \frac{1}{2^J\epsilon^J\pi^J}\mathrm{e}^{-\frac{J\,M^2}{2\,\epsilon^2}}\,\pi_0\left(C_M\right) \leq \tilde{Z}^y \tag{5.39}$$

Proof of Theorem 5.3.3

Let us begin with the definition of the Hellinger metric,

$$2d_{Hell}(\pi^y,\nu^y) = \int\left(\sqrt{\frac{\mathcal{L}^y(\mu,\sigma)}{Z}} - \sqrt{\frac{\tilde{\mathcal{L}}^y(\mu,\sigma)}{\tilde{L}}}\right)^2 d\pi_0(\mu,\sigma)$$

where $\mathcal{L}^y(\mu,\sigma)$ is the likelihood of the posterior π^y and $\tilde{\mathcal{L}}^y(\mu,\sigma)$ is the likelihood of the posterior ν^y, also Z and $\tilde{Z}$ are the normalisation constant of $\mathcal{L}^y(\mu,\sigma)$ and $\tilde{\mathcal{L}}^y(\mu,\sigma)$, respectively.

Having well-posedness of $L, \tilde{L}$ and $Z, \tilde{Z} \in (0,\infty)$, and, using $(\sqrt{x}-\sqrt{y})^2 \leq |x-y|$, using their definitions (5.7) and (5.21), we have that

$$\begin{aligned}
2d_{Hell}(\pi^y,\nu^y) &\leq \int \Big| \frac{\mathcal{L}^y(\mu,\sigma)}{Z} - \frac{\tilde{\mathcal{L}}^y(\mu,\sigma)}{\tilde{Z}} \Big| d\pi_0(\mu,\sigma) \\
&\leq \int \frac{\big|\mathcal{L}^y(\mu,\sigma) - \tilde{\mathcal{L}}^y(\mu,\sigma)\big|}{\tilde{Z}} + \big|\mathcal{L}^y(\mu,\sigma)\big| \Big|\frac{1}{Z} - \frac{1}{\tilde{Z}}\Big| d\pi_0(\mu,\sigma) \\
&= \int \frac{\big|\mathcal{L}^y(\mu,\sigma) - \tilde{\mathcal{L}}^y(\mu,\sigma)\big|}{\tilde{Z}} d\pi_0(\mu,\sigma) + \frac{\big|Z-\tilde{Z}\big|}{\tilde{Z}} \\
&\leq \frac{2}{\tilde{Z}} \int \big|\mathcal{L}^y(\mu,\sigma) - \tilde{\mathcal{L}}^y(\mu,\sigma)\big| d\pi_0(\mu,\sigma)
\end{aligned}$$

Let us estimate the difference of $\mathcal{L}^y(\mu,\sigma) - \tilde{\mathcal{L}}^y(\mu,\sigma)$, we have that

$$|\mathcal{L}^y(\mu,\sigma) - \tilde{\mathcal{L}}^y(\mu,\sigma)| = \left| \prod_{i=0}^{J-1} \tilde{p}^{(\mu,\sigma)}(y_{i+1},t_{i+1};y_i,t_i) - \prod_{i=0}^{J-1} \tilde{Q}^{(\mu,\sigma)}(y_{i+1},t_{i+1};y_i,t_i) \right| \tag{5.40}$$

For simplicity, let us define $p_i = \tilde{p}^{(\mu,\sigma)}(y_{i+1},t_{i+1};y_i,t_i)$, $q_i = \tilde{Q}^{(\mu,\sigma)}(y_{i+1},t_{i+1};y_i,t_i)$ and $\Delta t := \max_{i\in\{1,\dots J\}} |t_i - t_{i-1}|$. Then, for every $i \in \{0,\dots J\}$ we have that

$$\begin{aligned}
|p_i - q_i| &\leq \int \left| \left(p^{(\mu,\sigma)} - Q^{(\mu,\sigma)}\right) *_{t_{i+1}} f_{N(0,\epsilon^2)}(y_{i+1},t_{i+1};w,t_i) f_{N(0,\epsilon^2)}(y_i - w) \right| dw \\
&\leq \frac{C}{\sqrt{N}} \int \left(1+w^2\right) f_{N(0,\epsilon^2)}(y_i - w) dw = \frac{C(1+\epsilon^2+y_i^2)}{\sqrt{N}}
\end{aligned}$$

observe that the inequality holds due to Theorem 5.3.2 and we also note that C depends on C_1, C_2, C_3, Δt and ϵ.

We can easily see that the following inequality holds

$$\left| \prod_{i=0}^{J-1} p_i - \prod_{i=0}^{J-1} q_i \right| \leq \left[\prod_{j=0}^{J-2} (p_{j+1} \vee q_j) \right] \sum_{i=0}^{J-1} |p_i - q_i|$$

In addition, let us assume $N \geq \max\left\{ \Delta t \left(\frac{C_{\|\sigma\|_\infty} \mathrm{e}^{2C_{\|\mu'\|_\infty}\Delta t}}{\epsilon} \right)^2, \frac{C^2}{\epsilon^2} \right\}$. Then, we have that p_{i+1} is bounded

$$p_{i+1} \leq q_{i+1} + |p_{i+1} - q_{i+1}| \leq 2\max\left\{ \epsilon^{-1}, \frac{C_{\Delta t,\mu,\sigma,\epsilon}}{\sqrt{N}} \right\} \leq \frac{2}{\epsilon}$$

note that $q_{i+1} \leq \frac{1}{\epsilon}$ according to equation (5.38) in the Proof of Theorem 5.3.1. Also, since $N \geq \Delta t \left(\frac{C_{\|\sigma\|_\infty} \mathrm{e}^{2C_{\|\mu'\|_\infty}\Delta t}}{\epsilon} \right)^2$, we have that there exists a constant $\tilde{C}_{M,\epsilon,J}$ with $\tilde{Z}^y \geq \tilde{C}_{M,\epsilon,J}$, see equation (5.39) in the Proof of Theorem 5.3.1.

If we now apply the last four inequalities in (5.40), we obtain that the following difference is of order $O(\sqrt{N})$,

$$\frac{1}{\tilde{Z}^y}\left|\mathcal{L}^y(\mu,\sigma)-\tilde{\mathcal{L}}^y(\mu,\sigma)\right|\leq\frac{\tilde{C}}{\sqrt{N}}$$

with $\tilde{C}$ depends only on $J, M, \Delta t, \epsilon$ and $\pi_0(C_M)$. Therefore, we get the desired bound for the Hellinger inequality, i.e.

$$d_{Hell}(\pi^y,\nu^y)\leq\frac{C}{\sqrt{N}}$$

Proof of Theorem 5.3.2

Let us consider transition density function q as in (5.27) and define the following function

$$w(x,s):=p*_t f_{N(0,\epsilon^2)}(y,t;x,s)-q_{\xi,\tau}*_t f_{N(0,\epsilon^2)}(y,t;x,s) \tag{5.41}$$

for s within the time interval $[\tau,t]$. For simplicity, let us define the following two functions:

$$\begin{aligned} u(x,s)&=p*_t f_{N(0,\epsilon^2)}(y,t;x,s)\\ v(x,s)&=q_{\xi,\tau}*_t f_{N(0,\epsilon)}(y,t;x,s)\end{aligned} \tag{5.42}$$

Thus, the difference can be written as $w(x,s)=u(x,s)-v(x,s)$. Using the corresponding backward Kolmogorov equation for p and $q_{\xi,\tau}$, see in (5.6) and (5.28), and let us further denote by A and $\tilde{A}$ the differential operators appear on the right hand of their backward Kolmogorov equation, respectively. We then can easily see that u and v satisfy the following two equations $-\partial_s u(x,s)=Au(x,s)$ and $-\partial_s u(x,s)=\tilde{A}u(x,s)$, respectively. For seeing that we only need to observe that

$$\begin{aligned} -\partial_s v(x,s)&=\int f_{N(0,\epsilon^2)}(z)\left[-\partial_s q_{\xi,\tau}(y-z,t;x,s)\right]dz\\ &=\int f_{N(0,\epsilon^2)}(z)\left[\tilde{A}q_{\xi,\tau}(y-z,t;x,s)\right]dz=\tilde{A}v(x,s)\end{aligned} \tag{5.43}$$

Next, using the final conditions of p and $q_{\xi,\tau}$, see in (5.6) and (5.28), we get that u and v satisfy the same final condition $f_{N(0,\epsilon^2)}(y-x)$,

$$\lim_{s\nearrow t}v(x,s)=\lim_{s\nearrow t}q_{\xi,\tau}*_t f_{N(0,\epsilon^2)}(y,t;x,s)=\int f_{N(0,\epsilon^2)}(y-z)\delta_x(dz)=f_{N(0,\epsilon^2)}(y-x)$$

Therefore, we have that $w(x,s)$ satisfies the following differential problem with final conditions

$$\begin{aligned} -\partial_s w(x,s)&=Ap(x,s)-\tilde{A}q_{\xi,\tau}(x,s)=Aw(x,s)+g(x,s)\\ \lim_{s\nearrow t}w(y,t;x,s)&=0,\end{aligned} \tag{5.44}$$

where g is defined as it follows

$$g(x,s):=\frac{\sigma^2(x)-\tilde{\sigma}^2(s)}{2N}\partial_{xx}v(x,s)+(\mu(x)-\tilde{\mu}(x,s))\,\partial_x v(x,s) \tag{5.45}$$

Using the definition of v, see (5.42), the definition of $q_{\xi,\tau}$, see (5.27), and the same argument as in (5.37), we get that

$$v(y,t;x,s) = f_{N(\Phi(t;x,s),F_N(t;s))} * f_{N(0,\epsilon^2)}(y) = f_{N(\Phi(t;x,s),F_N(t;s)+\epsilon^2)}(y) \tag{5.46}$$

Then, let us evaluate the following derivatives of v. Using the form of u which is obtained in equation (5.46), we get that

$$\begin{aligned}\partial_s v(x,s) &= v(x,s)\left[-\partial_s \frac{(y-\Phi(t;x,s))^2}{2(F_N(t;s)+\epsilon^2)} - \frac{1}{2}\partial_s \log\left(F_N(t;s)+\epsilon^2\right)\right] \\ &= v(x,s)\left[\frac{y-\Phi(t;x,s)}{F_N(t;s)+\epsilon^2}\ \partial_s\Phi(t;x,s)\right. \\ &\qquad\left.+\left(\frac{(y-\Phi(t;x,s))^2}{2(F_N(t;s)+\epsilon^2)} - \frac{1}{2}\right)\partial_s \log\left(F_N(t;s)+\epsilon^2\right)\right]\end{aligned}$$

and

$$\partial_x v(x,s) = v(x,s)\left[\frac{y-\Phi(t;x,s)}{F_N(t;s)+\epsilon^2}\ \partial_x\Phi(t;x,s)\right]$$

Using equation (5.43), we get that

$$\begin{aligned}\frac{\tilde\sigma^2(s)}{2N}\partial_{xx}v(x,s) &= -\left[\partial_s + \tilde\mu(x,s)\ \partial_x\right]v(x,s) \\ &= -v(x,s)\Big[\frac{y-\Phi(t;x,s)}{F_N(t;s)+\epsilon^2}\ \left[\partial_s+\tilde\mu(x,s)\ \partial_x\right]\ \Phi(t;x,s) \\ &\qquad + \left(\frac{(y-\Phi(t;x,s))^2}{2(F_N(t;s)+\epsilon^2)} - \frac{1}{2}\right)\partial_s \log\left(F_N(t;s)+\epsilon^2\right)\Big]\end{aligned}$$

Then, it is easy to see that $[\partial_s + \tilde\mu(x,s)\ \partial_x]\ \Phi(t;x,s) = 0$, thus we have that

$$\frac{\tilde\sigma^2(s)}{2N}\partial_{xx}v(x,s) = v(x,s)\left(\frac{(y-\Phi(t;x,s))^2}{2(F_N(t;s)+\epsilon^2)} - \frac{1}{2}\right)\partial_s \log\left(\frac{1}{F_N(t;s)+\epsilon^2}\right)$$

Let us define the following two functions

$$\Sigma(x,s) = \frac{\sigma^2(x)-\tilde\sigma^2(s)}{\tilde\sigma^2(s)}$$

$$M(x,s) = \mu(x) - \tilde\mu(x,s)$$

Thus, the function g can be written as follows

$$\begin{aligned}g(x,s) = \Sigma(x,s)v(x,s)&\left(\frac{(y-\Phi(t;x,s))^2}{2(F_N(t;s)+\epsilon^2)} - \frac{1}{2}\right)\partial_s \log\left(\frac{1}{F_N(t;s)+\epsilon^2}\right) \\ &+ NM(x,s)\ v(x,s)\ \left[\frac{y-\Phi(t;x,s)}{NF_N(t,s)+N\epsilon^2}\ \partial_x\Phi(t;x,s)\right]\end{aligned}$$

Next, we see that $\partial_s \log\left(\frac{1}{F_N(t;s)+\epsilon^2}\right)$ is bounded, for every s within $[\tau,t]$,

$$0 \le \partial_s \log\left(\frac{1}{F_N(t;s)+\epsilon^2}\right) = \frac{[\sigma(\phi(s;\mu))\ \partial_x\Phi(t;x,s)]^2}{NF_N(t,s)+N\epsilon^2} \le \left(\frac{\|\sigma\|_\infty e^{\|\mu'\|_\infty(t-\tau)}}{\epsilon}\right)^2 =: C_1$$

in the last inequality, we use that $\partial_x \Phi\,(t;x,s) = \mathrm{e}^{\int_s^t \mu'(\,\phi_\tau(\lambda;\xi)\,)d\lambda} \leq e^{\|\mu'\|_\infty (t-\tau)}$. We also note that the constant C_1 depends on $t-\tau$, ϵ, $\|\mu\|_\infty$ and $\|\sigma\|_\infty$. Also, it is easy to obtain that for $x \in \mathbb{R}$, we have that $\mathrm{e}^{-|x|}\,|x| \leq \mathrm{e}^{-1}$ and $\mathrm{e}^{-x^2}\,|x| \leq \frac{1}{2\mathrm{e}^{1/4}}$. From the definition of v, see (5.46), we get that

$$v(x,s)\,|y-\Phi\,(t;x,s)| \leq \frac{1}{2\,\mathrm{e}^{1/4}\,\sqrt{\pi}} \qquad v(x,s)\frac{(y-\Phi\,(t;x,s))^2}{\sqrt{2\,(F_N(t;s)+\epsilon^2)}} \leq \frac{1}{\mathrm{e}\,\sqrt{\pi}} \tag{5.47}$$

and by definition of v, we have that $v(x,s) \leq \frac{1}{\sqrt{2\pi(F_N(t;s)+\epsilon^2)}}$. Hence, using the last three bounds and the bound for $\partial_s \log\left(\frac{1}{F_N(t;s)+\epsilon^2}\right)$, we have that

$$|g(x,s)| \leq C_1' \left(\frac{|\Sigma(x,s)|}{\sqrt{2\,(F_N(t;s)+\epsilon^2)}} + \frac{|M(x,s)|}{2\,(F_N(t;s)+\epsilon^2)} \right)$$

Since $F_N(t;s)$ is non-negative, we get that

$$|g(x,s)| \leq \tilde{C}_1\,(|\Sigma(x,s)| + |M(x,s)|)$$

with constant $\tilde{C}_1$ depends again on the same variables as C_1.

Using the lower bound of $\sigma_{\inf}$, combined with the Lipschitz condition of σ, we get that

$$|\Sigma(x,s)| = \frac{\left|\sigma^2(x) - \sigma^2\,(\,\phi_\tau\,(s;\mu)\,)\right|}{\sigma^2(\phi_\tau\,(s;\mu))} \leq C_\sigma\,|x-\phi_\tau\,(s;\mu)|$$

observe that the constant C_σ depends only on $\sigma_{\inf}$ and $\|\sigma\|_{0,1}$. Next, using the Lipschitz condition of μ', we get that

$$|M(x,s)| = \left|\mu(x) - \mu(\phi_\tau\,(s;\mu)) - \mu'(\phi_\tau\,(s;\mu))(x-\phi_\tau\,(s;\mu))\right| \leq C_\mu\,(x-\phi\,(s;\mu))^2$$

observe that the constant C_μ depends only on $\|\mu'\|_{0,1}$. Therefore, we have that

$$|g(x,s)| \leq C_2 \left(|x-\phi_\tau\,(s;\mu)| + (x-\phi_\tau\,(s;\mu))^2\right) \tag{5.48}$$

with C_2 depends on $t-\tau$, ϵ, $\sigma_{\inf}$, $\|\sigma\|_{0,1}$ and $\|\mu\|_{1,1}$. Hence, we have that g satisfies the following condition

$$|g(x,s)| \leq 2\,C_2 \sup_{s\in[\tau,t]} \phi_\tau^2(s;\mu)\ \left(1+x^2\right)$$

From the continuity of $\phi_\tau\,(s;\mu)$ on $s \in [\tau,t]$, we get that the above supremum is finite. Also, we have that $w(x,s)$ is bounded

$$\begin{aligned} |w(x,s)| &\leq \left|p *_t f_{N(0,\epsilon^2)}(y,t;x,s)\right| + \left|q_{\xi,\tau} *_t f_{N(0,\epsilon^2)}(y,t;x,s)\right| \\ &\leq \frac{1}{\epsilon\sqrt{2\pi}}\left(\left|\int p(z,t;x,s)dz\right| + \left|\int q_{\xi,\tau}(z,t;x,s)dz\right|\right) = \frac{2}{\epsilon\sqrt{2\pi}} \end{aligned}$$

the second line holds, since $f_{N(0,\epsilon^2)}(x) \leq \frac{1}{\epsilon\sqrt{2\pi}}$, and p, $q_{\xi,\tau}$ are density functions with respect to z. Using Feynman-Kac formula, see for instance Theorem 7.6 at page 366 in Karatzas and Shreve (2014), one can write w as it follows

$$|w(\xi,\tau)| = \left|\mathbb{E}\int_\tau^t g(X(s),s)ds\right| \leq C_2 \int_\tau^t \mathbb{E}\left|X(s) - \phi_\tau(s;\mu)\right| + \mathbb{E}\left(X(s) - \phi_\tau(s;\mu)\right)^2 ds$$

on the last inequality, we use that the right-hand side of inequality (5.48) is non-negative, thus Tonelli's theorem implies the interchange of the integral with the expectation. Let us apply Cauchy–Schwarz inequality, as it follows

$$|w(\xi,\tau)| \leq C_2 \int_\tau^t \sqrt{\mathbb{E}\left(X(s) - \phi_\tau(s;\mu)\right)^2} + \mathbb{E}\left(X(s) - \phi_\tau(s;\mu)\right)^2 ds$$

Then, Lemma 5.4.4 provide us with the following

$$|w(\xi,\tau)| \leq C_2 \int_\tau^t \sqrt{\frac{\tilde{C}}{N}(1+\xi^2)} + \frac{\tilde{C}}{N}(1+\xi^2)ds \leq \frac{\tilde{C}_2(1+\xi^2)}{\sqrt{N}}$$

with $\tilde{C}_2$ depends on $t-\tau$, ϵ, $\sigma_{\inf}$, $\|\sigma\|_{0,1}$ and $\|\mu\|_{1,1}$.

5.4.3 Lemmas for unbounded process

Lemma 5.4.1. *For a given $T>0$, we consider $X_0 \in C([0,T])$[2] and X_1 satisfies the SDE in equation (5.10), where $\mu \in C^1(\mathbb{R})$ and $\sigma \in C(\mathbb{R})$. Then, it holds that: for every $p \geq 1$*

$$\mathbb{E}\sup_{t\in[0,T]} |X_1(t)|^p \leq C, \qquad \text{for } T>0 \text{ and } p \geq 1$$

with C depending on $\|\mu'\|_\infty$, $\|\sigma\|_\infty$, p and T.

Proof. It is sufficient to show the lemma for $p \geq 2$. For $p \geq 2$ and $a, b \geq 0$, we have that $|x|^p$ is convex, thus we have: $(a+b)^p \leq 2^{p-1}(a^p + b^p)$. Next, we apply the integral form of equation (5.10) and the last inequality on the following

$$\mathbb{E}\sup_{t\in[0,T]} |X_1(s)|^p \leq 2^{p-1}\Big(\mathbb{E}\sup_{t\in[0,T]} \Big|\int_0^t X_1(q)\,\mu'(X_0(q))\,dq\Big|^p + \mathbb{E}\sup_{t\in[0,T]}\Big|\int_0^t \sigma(X_0(q))\,dW_q\Big|^p\Big) \tag{5.49}$$

Theorem 7.2 at p.40 in Mao (2007) provides us the following inequality,

$$\mathbb{E}\sup_{t\in[0,T]}\left|\int_0^t \sigma(X_0(q))\,dW_q\right|^p \leq K_{p,T}\int_0^T \mathbb{E}|\sigma(X_0(q))|^p dq \tag{5.50}$$

Let apply Hölder inequality for the first integral in (5.49) and also apply the last inequality

$$\mathbb{E}\sup_{t\in[0,T]} |X_1(t)|^p \leq K_{T,p}\left(\mathbb{E}\int_0^T |\mu'(X_0(q))\,X_1(q)|^p dq + \mathbb{E}\int_0^T |\sigma(X_0(q))|^p dq\right) \tag{5.51}$$

[2] It is not necessary that X_0 satisfies the ODE in (5.10), but the existence of the ODE implies $X_0 \in C([0,T])$

We have that $\mu'\left(X_0(q)\right)$ and $\sigma\left(X_0(q)\right)$ are bounded functions. We use Tonelli's theorem to interchange the following integrals,

$$\mathbb{E}\sup_{0\le s\le t}|X_1\left(s\right)|^p \le K_{T,p,\|\mu'\|_\infty}\int_0^t \mathbb{E}\sup_{0\le s\le q}|X_1(s)|^p dq + K_{T,p,\|\sigma\|_\infty}$$

Last, by applying Grönwall inequality, see for instance Lemma1.1 at page 30 in Freidlin and Wentzell (2012), we get the the desired result

$$\mathbb{E}\sup_{0\le s\le t}|X_1\left(s\right)|^p \le K_{T,p,\|\sigma\|_\infty} e^{K_{T,p,\|\mu'\|_\infty} t}$$

For $1 \le p < 2$ the result follows from an application of Hölder inequality. $\square$

Lemma 5.4.2. *For a given $T > 0$, we consider $\xi_0 \in C\left([0,T]\right)$ and ξ_1 satisfy the following two differential equations (5.16) and (5.18), respectively. Suppose also that $\mu \in C^1\left(\mathbb{R}\right)$ and $\sigma \in C\left(\mathbb{R}\right)$. Then, it holds that:*

$$\mathbb{E}\sup_{t\in[0,T]}|\xi_1(t)|^p \le C, \qquad \textit{for } T > 0 \textit{ and } p \ge 1$$

with C depending on $\|\mu'\|_\infty$, $\|\sigma\|_\infty$, p and T.

Proof. We first show the desired result for $p \ge 2$. As we discuss in Section 5.2.1, we can define Y_0 and $\tilde{Y}$ which satisfies equations (5.19), then ξ_0 and $\tilde{\xi}$ can be defined through Skorokhod map as it follows $\xi_0(t) = \Gamma(Y_0)(t)$ and $\tilde{\xi}(t) = \Gamma(\tilde{Y})(t)$. In addition, let us define $\xi_1^*(t) := \sqrt{N}\left(\tilde{Y}(t) - Y_0(t)\right)$, then we have that ξ_1^* satisfies the following equation

$$d\xi_1^*(t) = \sqrt{N}\mu'(\Gamma(Y_0)(t)\,)\left(\Gamma(\tilde{Y})(t) - \Gamma(Y_0)(t)\right)dt + \sigma(\Gamma(Y_0)(t)\,)dW_t$$

Since $p \ge 2$, for the evaluation of the expected value of the supremum of ξ_1^*, we can apply the same steps as that has been used in Lemma 5.4.1 to obtain an inequality analogous to inequality of (5.51). Then, we get that

$$\begin{aligned}\mathbb{E}\sup_{s\in[0,T]}|\xi_1^*(s)|^p \lesssim N^{p/2}\int_0^T \mathbb{E}\left|\mu'(\Gamma(Y_0)(q)\,)\left(\Gamma(\tilde{Y})(q) - \Gamma(Y_0)(q)\right)\right|^p ds \\ + \int_0^T |\sigma(\Gamma(Y_0)(q)\,)|^p\, ds\end{aligned}$$

and the constant which is implied by $\lesssim$, it only depends on p, T. From the continuity of μ', σ and $\Gamma(Y_0)(q)$, we have that both $\mu'(\Gamma(Y_0)(q)\,)$ and $\sigma(\Gamma(Y_0)(q)\,)$ are bounded on interval $q \in [0,T]$, thus we get that

$$\begin{aligned}\mathbb{E}\sup_{s\in[0,T]}\left|\sqrt{N}\left(\Gamma(\tilde{Y})(s) - \Gamma(Y_0)(s)\right)\right|^p &\le 2\mathbb{E}\sup_{s\in[0,T]}|\xi_1^*(s)|^p \\ &\lesssim \int_0^T \mathbb{E}\sup_{s\in[0,q]}\left|\sqrt{N}\left(\Gamma(\tilde{Y})(s) - \Gamma(Y_0)(s)\right)\right|^p dq + 1\end{aligned}$$

Observe that the first inequality is a property of the Skorokhod map, see for example part (i) in Lemma 3.2.1. We also note that the constant in the above inequality depends on the upper bound of μ', σ, p, T. At this point, it worths to observe that $\xi_1(q) = \sqrt{N}\left(\tilde{\xi}(q) - \xi_0(q)\right) := \sqrt{N}\left(\Gamma(\tilde{Y})(q) - \Gamma(Y_0)(q)\right)$, thus the Grönwall inequality implies that

$$\mathbb{E} \sup_{s\in[0,T]} |\xi_1(s)|^p \leq Ce^{CT}$$

For $1 \leq p < 2$ the result follows from an application of Hölder inequality. $\square$

Lemma 5.4.3. *Suppose $\mu, \tilde{\mu}$ are Lipschitz and the corresponding solutions $\phi_\tau(\cdot;\mu, y_\tau)$, $\tilde{\phi}_\tau(\cdot;\tilde{\mu}, y_\tau)$, similar as in Section 5.3, defined for times $t \geq \tau$. Then, it holds that for every $t \geq \tau$*

$$\sup_{s\in[\tau,t]} \left|\phi_\tau(s;\mu, y_\tau) - \tilde{\phi}_\tau(s;\tilde{\mu}, y_\tau)\right| \leq C\|\mu - \tilde{\mu}\|_\infty$$

with C depending on $\|\mu\|_{0,1}$ and T.

Proof. For simplicity instead of using the notation of $\phi_\tau(\cdot;\mu, y_\tau)$ mentioned in the statement of the Lemma, let us change it accordingly to the following:

$$y(s) = \phi_\tau(s;\mu, y_\tau) \qquad g(s) = \tilde{\phi}_\tau(s;\tilde{\mu}, y_\tau)$$

Using the differential form of y, we have that

$$\sup_{s\in[\tau,t]} |y(s)| \leq |y_\tau| + (t-\tau)\|\mu\|_\infty$$

Looking at the differential equation satisfied by the difference of y and g, we have that

$$\begin{aligned}\sup_{s\in[\tau,t]} \left|y(s) - g(s)\right| &\leq \int_\tau^t \left|\mu(y(q)) - \mu(g(q))\right| + \left|\mu(g(q)) - \tilde{\mu}(g(q))\right| dq \\ &\leq \int_\tau^t \|\mu\|_{0,1} \sup_{\tilde{q}\in[\tau,q]} \left|y(\tilde{q}) - g(\tilde{q})\right| + \|\mu - \tilde{\mu}\|_\infty dq\end{aligned}$$

Then, using Grönwall inequality, we obtain that desired inequality. $\square$

Lemma 5.4.4. *Suppose that $X^{(\mu,\sigma)}$ is the solution of equation (5.8) and $\phi_\tau(\cdot;\mu,\xi)$, or simply denoted by $\phi_\tau(\cdot;\xi)$, the solution of the deterministic ODE which appears in the system of equations in (5.10) with initial value ξ at time τ. In addition, let us assume that the coefficients μ, σ^2 are Lipschitz continuous. Then, for every $t \geq \tau$ there exists a constant C which depends on $\|\mu\|_{0,1}$, $\|\sigma^2\|_{0,1}$ and $t - \tau$ such that*

$$\sup_{s\in[\tau,t]} \mathbb{E}_{\xi,\tau}(X^{(\mu,\sigma)}(s) - \phi_\tau(\cdot;\mu,\xi))^2 \leq \frac{C\left(1+\xi^2\right)}{N}.$$

Proof. We observe that in this case, the Lipschitz condition of μ and σ^2 implies their linear growth condition, note that for every $x \in \mathbb{R}$ it holds the following

$$\begin{aligned} |\mu(x)| \leq K_\mu |x| + |\mu(0)| \leq \|\mu\|_{0,1}(1+|x|) \\ \left|\sigma^2(x)\right| \leq K_{\sigma^2}|x| + \sigma^2(0) \leq \|\sigma^2\|_{0,1}(1+|x|) \end{aligned} \tag{5.52}$$

The following argument implies that the solution X is bounded in the following way:

$$\begin{aligned} \mathbb{E}_{\xi,\tau} \sup_{q\in[\tau,s]} X_q^2 - \xi^2 \lesssim \mathbb{E}_{\xi,\tau} \sup_{q\in[\tau,s]} (\int_\tau^q \mu(X_q)dq)^2 + \frac{1}{N}\mathbb{E}_{\xi,\tau} \sup_{q\in[\tau,s]} (\int_\tau^q \sigma(X_q)dWq)^2 \\ \leq \mathbb{E}_{\xi,\tau} \int_\tau^s \mu(X_q)^2 dq(s-\tau) + \frac{1}{N}\mathbb{E}_{\xi,\tau} \int_\tau^s \sigma(X_q)^2 dq \\ \leq \mathbb{E}_{\xi,\tau} \int_\tau^s C(1+X_q^2)dq(s-\tau) + \frac{1}{N}\mathbb{E}_{\xi,\tau} \int_\tau^s C(1+|X_q|)dq \\ \leq \mathbb{E}_{\xi,\tau} \int_\tau^s C(1+X_q^2)dq(s-\tau) + \frac{1}{N}\mathbb{E}_{\xi,\tau} \int_\tau^s C(1+|X_q|)^2 dq \\ \lesssim \mathbb{E}_{\xi,\tau} \int_\tau^s C(1+X_q^2)dq(s-\tau) + \frac{1}{N}\mathbb{E}_{\xi,\tau} \int_\tau^s C(1+X_q^2)dq \\ = C(s-\tau+\frac{1}{N}) \int_\tau^s \mathbb{E}_{\xi,\tau} \sup_{\tilde{q}\in[\tau,q]} X_{\tilde{q}}^2 dq + C(s-\tau)(s-\tau+\frac{1}{N}) \end{aligned}$$

Grönwall's inequality implies that the following holds for every $s \in [\tau, t]$

$$\mathbb{E}_{\xi,\tau} \sup_{q\in[\tau,s]} X_q^2 \leq C(s-\tau)(s-\tau+1)\big(1+\xi^2\big)\mathrm{e}^{C(s-\tau+1)(s-\tau)}$$

observe that C depends on the linear growth of μ and σ^2. Using the boundedness of $\mathbb{E}_{\xi,\tau} \sup_{q\in[\tau,s]} X_q^2$, we get that

$$\begin{aligned} \mathbb{E}_{\xi,\tau}(X(s)-\phi_\tau(s;\xi))^2 = 2\,\mathbb{E}_{\xi,\tau} \int_\tau^s (X(q)-\phi_\tau(q;\xi))(\mu(X(q))-\mu(\phi_\tau(q;\xi)))dq + \frac{1}{N}\mathbb{E}_{\xi,\tau} \int_\tau^s \sigma(X(q))^2 dq \\ \leq 2K_\mu \int_\tau^s \mathbb{E}_{\xi,\tau}(X(q)-\phi_\tau(q;\xi))^2 dq + \frac{1}{N}K_{\sigma,lin.} \int_\tau^s \mathbb{E}_{\xi,\tau}(1+|X(q)|)dq \\ \lesssim \int_\tau^s \mathbb{E}_{\xi,\tau}(X(q)-\phi_\tau(q;\xi))^2 dq + \frac{1}{N} \int_\tau^s \mathbb{E}_{\xi,\tau}(1+X(q)^2)dq \\ \lesssim \int_\tau^s \mathbb{E}_{\xi,\tau}(X(q)-\phi_\tau(q;\xi))^2 dq + \frac{(s-\tau)}{N}(1+\mathbb{E}_{\xi,\tau} \sup_{q\in[\tau,s]} X(q)^2) \end{aligned}$$

Grönwall's inequality implies that the following holds for every $s \in [\tau, t]$

$$\mathbb{E}_{\xi,\tau}(X(s)-\phi_\tau(s;\xi))^2 \leq \tilde{C}\frac{(s-\tau)}{N}\mathrm{e}^{\tilde{C}(s-\tau)}\big(1+\xi^2\big)$$

observe that $\tilde{C}$ depends on the linear growth of μ and σ^2 and the Lipschitz constant of μ. Using inequality (5.52), we see that $\tilde{C}(s-\tau)\mathrm{e}^{\tilde{C}(s-\tau)}$ can be obtained as a function of $\|\mu\|_{0,1}$, $\|\sigma^2\|_{0,1}$ and $s-\tau$. □

Bibliography

Agapiou, S., Dashti, M., and Helin, T. (2021). Rates of contraction of posterior distributions based on p-exponential priors. *Bernoulli*, 27(3):1616–642. 83

Agapiou, S., Papaspiliopoulos, O., Sanz-Alonso, D., and Stuart, A. (2017). Importance sampling: Intrinsic dimension and computational cost. *Statistical Science*, pages 405–431. 3, 22

Anderson, R. F. and Orey, S. (1976). Small random perturbation of dynamical systems with reflecting boundary. *Nagoya Mathematical Journal*, 60:189–216. 41, 79

Bakry, D. and Émery, M. (1985). Diffusions hypercontractives. In Azéma, J. and Yor, M., editors, *Séminaire de Probabilités XIX 1983/84*, pages 177–206, Berlin, Heidelberg. Springer Berlin Heidelberg. 51

Blagoveshchenskii, Y. N. (1962). Diffusion processes depending on a small parameter. *Theory of Probability & Its Applications*, 7(2):130–146. 78

Bobkov, S. G. and Ledoux, M. (2000). From Brunn-Minkowski to Brascamp-lieb and to logarithmic Sobolev inequalities. *Geometric & Functional Analysis GAFA*, 10(5):1028–1052. 51

Bogachev, V. I., Krylov, N. V., Röckner, M., and Shaposhnikov, S. V. (2015). *Fokker-Planck-Kolmogorov Equations*, volume 207. American Mathematical Soc. 37

Croix, J.-C., Dashti, M., and Kiss, I. Z. (2020). Nonparametric Bayesian inference of discretely observed diffusions. 72, 74, 75, 76

Dashti, M. and Stuart, A. M. (2015). The Bayesian Approach to Inverse Problems. 11, 12, 14, 15, 16, 17, 18

Donoho, D. L. and Johnstone, I. M. (1998). Minimax Estimation via Wavelet Shrinkage. *The Annals of Statistics*, 26(3):879–921. 15

Doob, J. (1955). Martingales and one-dimensional diffusion. *Transactions of the American Mathematical Society*, 78(1):168–208. 29

Dynkin, E. B. (1965). *Markov processes*, pages 77–104. Springer Berlin Heidelberg, Berlin, Heidelberg. 36

Ethier, S. N. and Kurtz, T. G. (2009). *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons. 5

Evans, L. (2012). *An Introduction to Stochastic Differential Equations*. American Mathematical Society. 26

Evans, L. C. (1998). *Partial differential equations*, volume 19. Rhode Island, USA. 57

Feller, W. (1954a). Diffusion processes in one dimension. *Transactions of the American Mathematical Society*, 77(1):1–31. 29, 38

Feller, W. (1954b). The general diffusion operator and positivity preserving semi-groups in one dimension. *Annals of Mathematics*, pages 417–436. 29

Folland, G. (2013). *Real Analysis: Modern Techniques and Their Applications*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley. 12

Freidlin, M. I. and Wentzell, A. D. (2012). Random perturbations. In *Random Perturbations of Dynamical Systems*, pages 1–28. Springer. 86, 95

Gentil, I. (2003). The general optimal Lp-Euclidean logarithmic Sobolev inequality by Hamilton–Jacobi equations. *Journal of Functional Analysis*, 202(2):591–599. 51

Gross, L. (1975). Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083. 51

Holley, R. and Stroock, D. W. (1986). Logarithmic Sobolev inequalities and stochastic Ising models. 51

Jia, J., Peng, J., and Gao, J. (2016). Bayesian approach to inverse problems for functions with a variable-index Besov prior. *Inverse Problems*, 32(8):085006. 4, 52

Kallenberg, O. and Kallenberg, O. (1997). *Foundations of modern probability*, volume 2. Springer. 47

Karatzas, I. and Shreve, S. (2014). *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York. 27, 35, 94

Kolehmainen, V., Lassas, M., Niinimäki, K., and Siltanen, S. (2012). Sparsity-promoting Bayesian inversion. *Inverse Problems*, 28(2):025005. 4, 52

Kurtz, T. G. (1971). Limit Theorems for Sequences of Jump Markov Processes Approximating Ordinary Differential Processes. *Journal of Applied Probability*, 8(2):344–356. 5

LEDOUX, M. (1997). *CONCENTRATION OF MEASURE AND LOGARITHMIC SOBOLEV INEQUALITIES*. Institut de Mathematiques Universite Paul-Sabatier 31062, Toulouse (France). 45, 52, 53, 54, 58, 61, 62

Leporini, D. and Pesquet, J.-C. (2001). Bayesian wavelet denoising: Besov priors and non-Gaussian noises. *Signal Processing*, 81(1):55–67. Special section on Markov Chain Monte Carlo (MCMC) Methods for Signal Processing. 4, 52

Mao, X. (2007). *Stochastic Differential Equations and Applications*. Elsevier Science. 27, 33, 37, 39, 43, 85, 86, 87, 94

Mörters, P. and Peres, Y. (2010). *Brownian motion*, volume 30. Cambridge University Press. 39

Øksendal, B. (2010). *Stochastic Differential Equations: An Introduction with Applications*. Universitext. Springer Berlin Heidelberg. 37, 39

Otto, F. and Villani, C. (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis*, 173(2):361–400. 51

Pilipenko, A. (2014). *An introduction to stochastic differential equations with reflection*, volume 1. Universitätsverlag Potsdam. 30, 31, 33

Rantala, M., Vanska, S., Jarvenpaa, S., Kalke, M., Lassas, M., Moberg, J., and Siltanen, S. (2006). Wavelet-based reconstruction for limited-angle X-ray tomography. *IEEE Transactions on Medical Imaging*, 25(2):210–217. 4, 52

Renshaw, E. (2015). *Stochastic population processes: analysis, approximations, simulations*. OUP Oxford. 4

Skorokhod, A. V. (1961). Stochastic equations for diffusion processes in a bounded region. *Theory of Probability & Its Applications*, 6(3):264–274. 30, 31

Skorokhod, A. V. (1962). Stochastic equations for diffusion processes in a bounded region. ii. *Theory of Probability & Its Applications*, 7(1):3–23. 30

Skorokhod, A. V. (1982). *Studies in the theory of random processes*, volume 7021. Courier Dover Publications. 28

Talagrand, M. (1991). A new isoperimetric inequality for product measure and the tails of sums of independent random variables. *Geometric & Functional Analysis GAFA*, 1(2):211–223. 52

Tan Bui-Thanh, O. G. (2015). A scalable algorithm for MAP estimators in Bayesian inverse problems with Besov priors. *Inverse Problems & Imaging*, 9(1):27–53. 4, 52

Yamada, T. and Watanabe, S. (1971). On the uniqueness of solutions of stochastic differential equations. *Journal of Mathematics of Kyoto University*, 11(1):155–167. 28, 33

Zahm, O., Constantine, P., Prieur, C., and Marzouk, Y. (2019). Gradient-based dimension reduction of multivariate vector-valued functions. 49

Zahm, O., Cui, T., Law, K., Spantini, A., and Marzouk, Y. (2021). Certified dimension reduction in nonlinear Bayesian inverse problems. 3, 6, 45, 46, 47, 48, 49, 51, 52, 53