Bayesian Non-linear Statistical Inverse Problems. [1]

[This version: March 14, 2023]

[1] Richard Nickl, nickl@maths.cam.ac.uk; Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge, UK.

PREFACE

These notes are based on graduate ('Nachdiplom') lectures given in spring term 2022 at the Department of Mathematics, ETH Zürich, Switzerland. I am very grateful to the Forschungsinstitut für Mathematik (FIM) at ETH Zürich for hosting me, as well as to Afonso Bandeira and Sara van der Geer for their hospitality during my visit.

Chapters 1 and 2 develop the framework of Bayesian Gaussian process methods in non-linear random design regression models and give sufficient conditions to obtain global convergence guarantees for posterior measures in PDE-type inverse problems. Chapters 3 to 5 develop the local theory about fluctuations and shape of posterior measures in high dimensions, and touch on the related issue of convergence properties of gradient based MCMC algorithms. The reader will require background in real analysis, measure-theoretic probability and stochastic convergence theory – one may consult [47] or similar texts. We will frequently use mathematical techniques from high-dimensional statistics and probability as developed in [61]. Relevant material from elliptic partial differential equations, functional analysis and stochastic calculus will be reviewed in the appendix.

I would like to thank the many colleagues whose feedback and input over the past years have helped shape my understanding of this material. Let me mention Francis Bach, Afonso Bandeira, Ismaël Castillo, Victor Chernozhukov, Marc Hoffmann, Ankur Moitra, Francois Monard, Kolyan Ray, Sebastian Reich, Markus Reiß, Philippe Rigollet, Judith Rousseau, Christoph Schwab, Vladimir Spokoiny, Botond Szabo, Andrew Stuart, Edriss Titi, Sasha Tsybakov, Sara van de Geer, Aad van der Vaart, Harry van Zanten and Martin Wainwright; as well as current, former and visiting members of my research group, among them Kweku Abraham, Randolf Altmeyer, Jan Bohr, Matteo Giordano, Hanne Kekkonen, Aksel K. Rasmussen, and Sven Wang. I would particularly like to thank Gabriel P. Paternain for the many inspiring discussions about non-linear inverse problems over the past years – these notes owe much to them.

RN

Cambridge, March 2023.

Contents

Die Schriftsteller
auch wenn sie Wissenschaftler sind
sind Übertreibungsspezialisten

T.B.

Chapter 1

Non-linear statistical inverse problems

The study of inverse problems forms an active field at the interface of applied and pure mathematics as well as the statistical and physical sciences. Prototypical examples include parameter identification in partial differential equations (PDEs) as well as various tomography and data assimilation tasks. The Bayesian approach to such problems has seen substantial activity in the last decade after seminal work by Andrew Stuart [119] – the recent references [36], [81, 109], [40], [25], [5] among many others survey the relevance of such inference techniques in various areas of applied mathematics. Some of these ideas can be traced back as far as Henri Poincaré – see the discussion in Persi Diaconis' early contribution [43], which is itself a relevant predecessor to the use of Bayesian thinking in this field.

The recent surge of interest stems from the inherent ability of the Bayesian approach to automatically and simultaneously address a variety of challenges of contemporary data science. These include algorithmic feasibility even for non-convex problems in high- or infinite-dimensional parameter spaces (via Markov chain Monte Carlo methods) as well as the provision of uncertainty quantification methods and 'error bars' for algorithmic outputs. A scientist using such methodology needs to specify a numerically tractable likelihood-function modelling the statistical measurement process and a reasonable prior distribution (such as a Gaussian random process) on the parameter, and in return obtains a posterior distribution that in principle can be used to solve all inferential tasks.

As much as the Bayesian 'package' is attractive in applications, a framework providing rigorous statistical and algorithmic *guarantees* for such methods – such as a convergence analysis in the large sample size scenario – has been developed only recently. For linear inverse problems the Bayesian approach is fairly well understood (e.g., [77, 108] and also [62, 65, 90] and references therein) and can be related to the well established 'regularisation' literature [52], [74], [17], where a convex penalised least squares fit functional is minimised to reconstruct the parameter of interest. But for *non-linear* problems Bayesian methods are genuinely distinct from optimisation based approaches and hence the theory requires different ideas. The purpose of these notes is to lay out some of the main mathematical mechanisms underpinning a body of recent theoretical work [94], [91, 92], [102], [1, 23, 63, 76, 101] on this subject. It builds on previous work on the understanding of Bayesian procedures in general high- and infinite-dimensional models by Aad van der Vaart and co-authors (e.g., [129] and the monograph [59]).

A common setting involves a non-linear 'forward map' $\mathscr{G}$ defined on some infinite-dimensional 'parameter' space Θ and taking values in some Hilbert space of functions or operators. We can think of $\mathscr{G}(\theta)$ describing the solution of a PDE in dependence of an unknown function (or 'coefficient') f_θ modelled by some parameter θ. The inverse problem is to recover θ from 'data' $\mathscr{G}(\theta)$. Even if $\theta \mapsto \mathscr{G}(\theta)$ is proved to be injective so that an inverse $\mathscr{G}^{-1}$ exists in principle, solution maps $\mathscr{G}$ of PDEs are typically 'smoothing' (compactifying) which in infinite-dimensions means that $\mathscr{G}^{-1}$ may not be continuous in relevant topologies. Also, since $\mathscr{G}$ is non-linear, we typically do not have an explicit representation of the inverse map and analytical formulæ reconstructing θ from $\mathscr{G}$ (such as, e.g., for Fourier, Laplace and Radon transforms available in linear settings) are out of reach. And the problem becomes even less tractable if we explicitly acknowledge discretisation and statistical measurement error. Nevertheless we will show that a principled Bayesian 'likelihood' approach may overcome such obstacles and, under certain conditions, provide provably valid algorithmic solutions for non-linear inverse problems.

1.1 Model examples

Non-linear forward maps $\mathscr{G}$ arise in a variety of settings with spectral inverse problems, X-ray transforms, boundary rigidity problems, data assimiliation and more generally in parameter identification problems with PDEs – see [75], [119], [74], [81, 109], [69] and also the recent monograph [105] by Gabriel Paternain, Mikko Salo and Gunther Uhlmann. We first describe some classical examples.

1.1.1 Boundary measurements in tomography

The Calderón problem

Let $\mathcal{X} \subset \mathbb{R}^d, d \geq 2$, be a bounded domain with smooth boundary $\partial\mathcal{X}$ and $\theta : \mathcal{X} \to [\theta_{min}, \infty)$ a differentiable map. Consider unique solutions $u = u_{\theta,h}$ to the PDE

$$\begin{aligned} \nabla \cdot (\theta \nabla u) &= 0 \quad \text{in } \mathcal{X}, \\ u &= h \quad \text{on } \partial\mathcal{X}, \end{aligned}$$

where $h : \partial\mathcal{X} \to \mathbb{C}$ prescribes some boundary values. (Existence of such solutions is discussed in more detail after (1.2) below.) If we cannot measure u inside of $\mathcal{X}$, we may record boundary information of $u_{\theta,h}$ for sufficiently many distinct h. Specifically one can measure the *Neumann (boundary) data* as

$$\Lambda_\theta(h) := \frac{\partial u_{\theta,h}}{\partial \nu}\Big|_{\partial\mathcal{X}},$$

where $\partial/\partial\nu$ denotes the outward normal derivative on $\partial\mathcal{X}$ (to be understood in a trace sense). The inverse problem here is to recover θ from knowledge of the boundary operator $\mathscr{G}(\theta) = \Lambda_\theta$ only – this is known as the Calderón problem , which has fundamental applications in *electric impedance tomography*, where θ models a conductivity inside of $\mathcal{X}$ to be recovered from non-invasive measurements. See [124] for an overview of the theory and applications of this problem (first studied by Alberto Calderón [26]), and [93, 120] for landmark injectivity theorems for $\mathscr{G}$.

Non-Abelian X-ray transforms

Let $M \subset \mathbb{R}^2$ be the closed unit disk with boundary ∂M. Consider lines in the plane (i.e. geodesics) parametrised by $\gamma(t) = x + tv$, where $x \in \mathbb{R}^2$ and $v \in S^1$. We further introduce the influx boundary as

$$\partial_+ SM = \{(x,v) \in \partial M \times S^1 : \; x \cdot v \leq 0\},$$

where $\cdot$ is the standard dot product in the plane. If we take $(x,v) \in \partial_+ SM$, then the line $\gamma(t) = x + tv$ will exit the disk in finite time $\tau_\gamma = \tau(x,v) := -2\, x \cdot v$.

Let $\theta : M \to \mathbb{C}^{n\times n}$ be a continuous matrix field. Given a geodesic $\gamma : [0, \tau_\gamma] \to M$ with endpoints $\gamma(0), \gamma(\tau_\gamma) \in \partial M$, we consider the matrix ODE

$$\dot{U} + \theta(\gamma(t))U = 0, \quad U(\tau_\gamma) = Id.$$

We define the boundary scattering data of θ on γ to be $C_\theta(\gamma) := U(0)$. This problem, backward in time for convention here, is well-posed and leads to a unique definition of $U(0)$, containing information about θ along the geodesic γ. Note that when θ is scalar ($n = 1$), we obtain

$$\log U(0) = \int_0^{\tau_\gamma} \theta(\gamma(t))dt,$$

which is the classical X-ray/Radon transform [107] of θ along the ray γ. Considering the collection of all such data makes up the *non-Abelian X-ray transform* of θ, viewed here as the non-linear map

$$\mathscr{G}(\theta) \equiv C_\theta \colon \partial_+ SM \to \mathbb{C}^{n\times n}, \tag{1.1}$$

and the goal is to recover θ from C_θ. We refer to [103, 104] and references therein for injectivity theorems for $\mathscr{G}$. A concrete physical application arises when θ takes values in the Lie algebra $so(n)$ of skew-symmetric matrices associated to the special orthogonal group $SO(n)$. In this case the scattering data C_θ maps into $SO(n)$ and for $n = 3$ this is relevant in *neutron spin tomography*, see [68, 115] and also Figure 1.1 below.

1.1.2 Parameter identification for elliptic PDEs

The two preceding examples are within the scope of the general theory to be developed here – see Remark 2.3.4 below and the notes to subsequent chapters for more details. But in these lecture notes we will focus exclusively on two non-linear model examples with *elliptic* PDEs that arise from steady state measurements of heat-type equations. Their relative analytical simplicity allows us to give a largely self-contained treatment of the material, yet at the same time these examples provide a 'pedagogical' template for more complicated settings.

To this end, let us introduce our elliptic base examples. Let $\mathcal{X}$ be a bounded domain in $\mathbb{R}^d$ with smooth boundary $\partial\mathcal{X}$. For a linear elliptic differential operator $\mathscr{L}_f$ indexed by some function $f : \mathcal{X} \to \mathbb{R}$ we consider solutions $u = u_f$

$$\begin{aligned} \mathscr{L}_f u &= g \quad \text{on } \mathcal{X}, \\ u &= h \quad \text{on } \partial\mathcal{X}, \end{aligned} \tag{1.2}$$

where g is a given smooth source function and h prescribes smooth boundary values. For the examples of $\mathscr{L}_f$ studied in these notes, unique solutions u_f to the boundary value problem (1.2) exist by standard results for linear elliptic PDEs (see after (6.10) in the appendix). We will later parameterise $f = f_\theta$ by a function θ varying in a linear space but suppress this in the notation for now.

If $\nabla\cdot, \nabla$ and Δ denote the divergence, gradient and Laplace operations respectively, our first example will be the divergence form operator $\mathscr{L}_f$ described by the action on sufficiently regular functions $u : \mathcal{X} \to \mathbb{R}$ via

$$\mathscr{L}_f u = \nabla \cdot (f \nabla u) = \nabla f \cdot \nabla u + f \Delta u, \quad f \geq f_{min} > 0. \tag{1.3}$$

Here f models a real-valued *diffusivity* (or conductivity) of the medium $\mathcal{X}$ and as explained in more detail in the next subsection, (1.2) describes 'steady state' solutions of a *diffusion equation.* Determining f from solutions u_f to (1.2) has various applications in physics and engineering, for instance in the context of 'groundwater flow' it is sometimes called *Darcy's problem* (see Section 3.7 in [119]). Unlike in the Calderón problem from earlier which deals with the same PDE, we will assume that we can measure u_f *throughout* $\mathcal{X}$ and not just at $\partial\mathcal{X}$ – see Remark 2.3.4 below for a discussion of the resulting differences.

The second example we shall examine closely is a *Schrödinger operator* $\mathscr{L}_f$ given by the action

$$\mathscr{L}_f u = \frac{\Delta u}{2} - fu, \quad f \geq 0, \tag{1.4}$$

where the absorption potential f models an attenuation in the diffusivity (e.g., a cooling effect in the heat flow) – in this case we speak of (1.2) as the (steady state) *Schrödinger equation* – even though we do not consider the more complicated *complex* valued version of (1.2) relevant in quantum physics. Determining f from solutions u_f to (1.2) for such $\mathscr{L}_f$ is a problem appearing for instance in photo-acoustic tomography [8, 9]. In (1.4), Δ could be replaced by any given (known) elliptic second order differential operator but we stick to the standard Laplacian for (mostly notational) simplicity. The Schrödinger model is in a certain sense atypical for most non-linear inverse problems as it gives a direct inversion formula for $f = \Delta u/2u$ (provided u does not vanish on $\mathcal{X}$). But it will be very useful to succinctly explain several mathematical ideas in these notes related to the 'local curvature' of typical non-linear forward maps $f \mapsto u_f$ encountered in the theory. It also provides a relevant template for general operators of the form $\mathscr{L}_f = \mathscr{D} + f$ with f an order-zero perturbation of a given (not necessarily elliptic) differential operator $\mathscr{D}$. One example would be the one underlying (1.1), where $\mathscr{D}$ arises from the geodesic vector field. See the notes to this section for more discussion.

We note that while the operator $\mathscr{L}_f$ in the preceding two examples is linear, the forward map $\mathscr{G}$ arising from $f \mapsto u_f$ is not – see Ex. 1.4.1. It will however be injective (under conditions to be detailed below), and the problem of inferring f from u_f is then a well-posed non-linear inverse problem.

From heat to steady state equation

As we will later consider noisy 'real world' measurements of 'steady state' solutions u_f of (1.2), we digress briefly and explain a way to acquire approximate measurements of u_f from observations of the process of *diffusion* over a time interval $(0, T]$. Consider solutions $\{u = u_f(x,t) : x \in \mathcal{X}, t \in (0,T]\}$ of the time evolution equation

$$\begin{aligned} \frac{\partial u}{\partial t} &= \mathscr{L}_f u \quad \text{on } \mathcal{X} \times (0,T] \\ u(\cdot, t) &= 0 \quad \text{on } \partial\mathcal{X}\ \forall t \in (0,T] \\ u(\cdot, 0) &= g \quad \text{on } \mathcal{X}, \end{aligned} \tag{1.5}$$

where g is some initial condition. It is well known (see Subsection 6.1.2 for details) that the elliptic operators $\mathscr{L}_f$ from (1.3), (1.4) equipped with Dirichlet boundary conditions (i.e., on the space H_0^1 from Subsection 6.1.1) have a spectral representation in terms of eigen-pairs

$$(\lambda_{j,f}, e_{j,f}) \in (0,\infty) \times H_0^1(\mathcal{X}), j \in \mathbb{N},$$

of the form

$$\mathscr{L}_f = -\sum_{j \geq 1} \lambda_{j,f} e_{j,f} \langle e_{j,f}, \cdot \rangle_{L^2} \tag{1.6}$$

where $\langle \cdot, \cdot \rangle_{L^2}$ is the standard inner product of the Hilbert space $L^2(\mathcal{X}) = L^2(\mathcal{X}, dx)$ with dx Lebesgue measure. The unique solutions to the heat equation (1.5) with initial condition g then have the well known representation

$$u_f(\cdot, t)(g) = \sum_j e^{-t\lambda_{j,f}} e_{j,f} \langle e_{j,f}, g \rangle_{L^2}, \quad t \geq 0.$$

Formally integrating term by term

$$\int_a^b e^{-\lambda_{j,f} s} ds = -\frac{1}{\lambda_{j,f}} \left[e^{-\lambda_{j,f} b} - e^{-\lambda_{j,f} a} \right], \quad a < b, \quad j \in \mathbb{N}, \tag{1.7}$$

and taking limits $a \to 0, b \to \infty$, we see that time averages of $u_f(\cdot, s)(g)$

$$-\int_0^\infty u_f(\cdot, s)(g) ds = -\sum_{j \geq 1} \lambda_{j,f}^{-1} e_{j,f} \langle e_{j,f}, g \rangle_{L^2} = \mathscr{L}_f^{-1}(g) \tag{1.8}$$

provide the solution to (1.2) with $h = 0$ (Dirichlet boundary conditions). So a Riemann sum approximation

$$-\sum_l u_f(\cdot, s_l)(g)[s_l - s_{l-1}] \tag{1.9}$$

where $u_f(\cdot, s_l)(g)$ are measurements of the solution of (1.5) at discrete equally spaced times $s_l \in (0, \infty)$, will give an approximate measurement of the 'steady state' solution u_f of (1.2). [Boundary values different from $h = 0$ can be considered likewise by modifying the base heat equation (1.5) accordingly.]

1.1.3 Data assimilation

While steady state approximations to time evolution equations such as those just described are commonly used in applications, sometimes modelling time horizon dynamics is the explicit task of statistical inference – for instance, in 'data assimilation' problems [37, 81, 109]. A prototypical example arising in atmospherical sciences and fluid dynamics is tracking a velocity field describing the solutions to the non-linear (incompressible) *Navier-Stokes equations* with unknown initial condition. If $\nu > 0$ denotes a scalar viscosity parameter, we can consider two-dimensional vector fields $u = u_\theta = (u_\theta(t,x) : t \in (0,T], x \in \Omega), \Omega \subset \mathbb{R}^2$, solving the system of non-linear partial differential equations given by

$$\begin{aligned} \frac{\partial}{\partial t}u - \nu\Delta u + u \cdot \nabla u + \nabla p = 0 \quad & \text{on } \Omega \times (0,T] \\ \nabla \cdot u = 0 \quad & \text{on } \Omega \times (0,T] \\ u = 0 \quad & \text{on } \partial\Omega \\ u(0,\cdot) = u_0 \quad & \text{on } \Omega, \end{aligned}$$

where $u(0)$ is a divergence free initial condition and ∇p a pressure term, see [35] for details. One regards the initial condition $u(0) \equiv \theta$ as the unknown parameter and the solutions $\mathscr{G}(\theta) = u_\theta$ measured at discrete points (t_i, X_i) in time and space as the 'forward' data. The aim is to infer θ and then also all subsequent states $(u_\theta(t,x) : t > 0, x \in \Omega)$ of the system.

In the above example the non-linearity of $\mathscr{G}$ arises from the underlying PDE – had we used the standard heat equation (1.5) instead, the resulting solution map $\theta \mapsto u_\theta$ would have been linear. With heat-equation type data arising from the generator $\mathscr{L}_f$, non-linearities still arise for the coefficient to solution maps $f \mapsto u_f(t,\cdot)$. For some observational settings, steady state approximations can be infeasible, specifically when measurements are available only at low 'observation frequency' in time t. In this case, knowledge of the Markovian time evolution structure should be exploited to solve the inverse problem. We will not prove rigorous theorems about such data assimilation problems here, but the key ideas and techniques apply to these settings as well – we refer to [64, 95, 97, 100] for some first results in this direction.

1.2 Bayesian inference in non-linear regression models

1.2.1 The forward map $\mathscr{G}$

We now introduce a notational framework designed to accommodate a large variety of examples of non-linear inverse problems. Let $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Z}, \mathcal{B})$ be measurable spaces equipped with measures λ, ζ, respectively. We assume that λ is a probability measure and that ζ is a finite measure. Let further V, W be vector spaces of fixed finite dimensions $p_V, p_W \in \mathbb{N}$, with inner products $\langle\cdot,\cdot\rangle_W, \langle\cdot,\cdot\rangle_V$ and norms $|\cdot|_W, |\cdot|_V$, respectively. Then

$$L^\infty(\mathcal{X}), \quad L^2(\mathcal{X}) = L^2_\lambda(\mathcal{X}, V) \quad \text{and} \quad L^\infty(\mathcal{Z}), \quad L^2(\mathcal{Z}) = L^2_\zeta(\mathcal{Z}, W)$$

denote the bounded measurable, and λ- or ζ- square integrable, V- or W-valued vector fields defined on $\mathcal{X}, \mathcal{Z}$, respectively. Denote by

$$\|\cdot\|_{L^2_\zeta(\mathcal{Z})}, \|\cdot\|_{L^2_\lambda(\mathcal{X})}, \ \langle\cdot,\cdot\rangle_{L^2(\mathcal{Z})}, \langle\cdot,\cdot\rangle_{L^2(\mathcal{X})}$$

the usual L^2-norms and Hilbert space inner products on these spaces. We write generically $\|\cdot\|_\infty$ for the supremum norm of a function over its domain.

We will consider parameter spaces Θ that are (Borel-measurable) subsets of $L^2(\mathcal{Z}, W)$ – later these will often have to be *linear* spaces so that Gaussian measures can be defined on them, but for now this is not necessary. On Θ, measurable 'forward maps'

$$\theta \mapsto \mathscr{G}(\theta), \qquad \mathscr{G} : \Theta \to L^2_\lambda(\mathcal{X}, V), \tag{1.10}$$

are defined. For the PDEs introduced in Subsection 1.1.2 we will regard the coefficient $f = f_\theta$ as being parameterised by $\theta \in \Theta$, and $\mathscr{G}$ will be the solution map $\theta \mapsto u_{f_\theta}$ of (1.2). In this case we have $V = W = \mathbb{R}$ and $\mathcal{X} = \mathcal{Z}$, but in many other inverse problems such as those from Subsections 1.1.1, 1.1.3 (or see [75, 105]), a more flexible choice of $V, W, \mathcal{X}, \mathcal{Z}$ is of interest which is why we present the general theory in this way.

1.2.2 A random design regression model with normal errors

Real-world measurements in inverse problems arising from forward data $\mathscr{G}(\theta)$ in (1.10) are discrete and subject to observational noise. For instance with solutions u_f of (1.2) one typically discretises $\mathcal{X}$ into a finite set $X_1, \dots, X_N$ for which measurements $u_f(X_i)$ are taken. We regard the X_i's as chosen *at random* from $\mathcal{X}$ according to the probability distribution λ. This viewpoint of 'probabilistic numerics' (see the notes to this section for discussion) combines naturally with our second source of randomness – the additive measurement error ε_i occurring when performing a measurement of $\mathscr{G}(\theta)(X_i)$. We will follow Gauss [57] and assume that ε_i follows a normal (Gaussian) distribution, which has a solid probabilistic foundation – for instance if our observations arise from time averages (1.9) of solutions u_f of the underlying heat equation, then each ε_i itself is already a cumulative sum of independent measurement errors, which by the central limit theorem of probability will be approximately normally distributed. We note, however, that the assumption that the ε_i be Gaussian is *not* necessary, see the notes to this chapter.

To fix ideas suppose that observations arise as the jointly independent and identically distributed (i.i.d.) random vectors $(Y_i, X_i)_{i=1}^N$ of the form

$$Y_i = \mathscr{G}(\theta)(X_i) + \varepsilon_i, \quad \varepsilon_i \sim^{i.i.d} N(0, I_V), \quad i = 1, \dots, N, \tag{1.11}$$

where $\mathscr{G}$ is as in (1.10), and the X_i's are random i.i.d. covariates drawn from law λ on $\mathcal{X}$, independent of the noise ε_i. We assume that the the inner product of V is chosen such that the covariance matrix I_V of each noise vector $\varepsilon_i \in V$ is diagonal.

The joint law of the random variables $(Y_i, X_i)_{i=1}^N$ in (1.11) defines a product probability measure on $(V \times \mathcal{X})^N$, and it will be denoted by $P_\theta^N = \otimes_{i=1}^N P_\theta^i$, where we note $P_\theta^i = P_\theta^1$ for all i. We write $P_\theta = P_\theta^i$ for the law of a 'generic copy' (Y, X), which has probability density (Radon-Nikodym derivative)

$$\frac{dP_\theta}{d\mu}(y,x) \equiv p_\theta(y,x) = \frac{1}{(2\pi)^{p_V/2}} \exp\Big\{ -\frac{1}{2}|y - \mathscr{G}(\theta)(x)|_V^2 \Big\}, \quad y \in V, x \in \mathcal{X}, \tag{1.12}$$

for dominating measure $d\mu = dy \times d\lambda$, with dy Lebesgue measure on V.

The infinite product probability measure $\otimes_{i=1}^\infty P_\theta$ describing the law of all possible infinite sequences of observations (in $(V \times \mathcal{X})^{\mathbb{N}}$) will be denoted by $P_\theta^{\mathbb{N}}$. The usual stochastic O_P, o_P notation will be used throughout for $P = P_\theta^N$ or equivalently $P_\theta^{\mathbb{N}}$ – see [127] for the standard theory of stochastic convergence. The expectation operator of $P_\theta^{\mathbb{N}}$ is denoted by E_θ and we sometimes write E_λ for the expectation under the X_i's only, while E_ε denotes expectation under the noise variables ε_i only. We also write shorthand

$$D_N := \{Y_1, \dots, Y_N, X_1, \dots, X_N\}, \qquad N \in \mathbb{N}, \tag{1.13}$$

for the full data vector.

1.2.3 The Bayesian prior and posterior distribution

We now let Θ be a measurable subset of $L^2_\zeta(\mathcal{Z}, W)$ with trace Borel-σ-algebra $\mathcal{F}$ – in fact all that follows in this subsection works for parameter spaces Θ which are Polish spaces equipped with their Borel σ-algebra $\mathcal{F}$. Let Π be a probability measure on $(\Theta, \mathcal{F})$ that we will often just call the *'prior'*. On the product space $\Theta \times (V \times \mathcal{X})$ with tensor Borel σ-algebra $\mathcal{F} \otimes (\mathcal{B}_V \otimes \mathcal{A})$ we can consider a probability measure Q given by the density

$$dQ(\theta, (y,x)) = p_\theta(y,x) d\mu(y,x) d\Pi(\theta), \quad y \in V, x \in \mathcal{X}, \theta \in \Theta,$$

where p_θ is as in (1.12), and where we assume that $(\theta, x) \mapsto \mathscr{G}(\theta)(x)$ is jointly $\mathcal{F} \otimes \mathcal{A} - \mathcal{B}_{\mathbb{R}}$ measurable. This space formally supports the Bayesian model: the random variables $(Y, X)|\theta$ have conditional μ-densities

$$\frac{p_\theta(y,x) d\Pi(\theta)}{\int_{V \times \mathcal{X}} p_\theta(y,x) d\mu(y,x) d\Pi(\theta)} = p_\theta(y,x),$$

as in (1.12), whereas $\theta|(Y, X)$ has 'posterior' density

$$\frac{p_\theta(Y,X) d\Pi(\theta)}{\int_\Theta p_\theta(Y,X) d\Pi(\theta)}.$$

Then for an i.i.d. sample from this model we have likewise

$$(Y_i, X_i)_{i=1}^N | \theta \sim^{iid} P_\theta^N$$

with *posterior distribution* $\Pi(\cdot|D_N) = \Pi(\cdot|(Y_i, X_i)_{i=1}^N)$ of $\theta|(Y_i, X_i)_{i=1}^N$ on Θ now given by the ratio

$$d\Pi(\theta|D_N) = \frac{\prod_{i=1}^N p_\theta(Y_i, X_i) d\Pi(\theta)}{\int_\Theta \prod_{i=1}^N p_\theta(Y_i, X_i) d\Pi(\theta)} = \frac{e^{\ell_N(\theta)} d\Pi(\theta)}{\int_\Theta e^{\ell_N(\theta)} d\Pi(\theta)}, \ \theta \in \Theta, \tag{1.14}$$

where we introduce the notation

$$\ell_N(\theta) = \sum_{i \le N} \ell_i(\theta), \quad \text{for } \ell_i(\theta) = -\frac{1}{2}|Y_i - \mathscr{G}(\theta)(X_i)|_V^2, \tag{1.15}$$

which is, up to additive constants, the log-likelihood function $\log dP_\theta^N(D_N)$ of the data D_N. At this point the preceding ratios are well defined at least Q-almost surely because they are conditional distributions/densities on the product space $\Theta \times (V \times \mathcal{X})$. A second 'frequentist' way to make sense of these ratios is discussed before Lemma 1.3.3 below. See also Chapter 1 in [59] for more discussion of the latter aspect.

1.2.4 Posterior computation by MCMC

A Bayesian statistician will base its inferences about θ on the posterior distribution $\Pi(\cdot|D_N)$ from (1.14). In our context where $\mathscr{G}$ is *non-linear*, the expression in (1.14) remains abstract and in absence of analytical formulæ it is unclear how we can extract practical statistical information from it, specifically if we wish to avoid computation of the normalising factor which would involve the evaluation of a possibly intractable integral over Θ. It is here where Markov chain Monte Carlo (MCMC) methods enter the stage. The main idea behind MCMC methods is to generate a Markov chain that has the posterior distribution as invariant measure. This can often be done in a way that requires at each step a single evaluation of the likelihood function $\ell_N(\theta)$ from (1.15), which in turn involves an evaluation of $\mathscr{G}$ rather than of its typically much more complicated inverse $\mathscr{G}^{-1}$, and also bypasses calculation of the normalising factor in (1.14). If such a Markov chain (ϑ_k) is ergodic and 'mixes well' we can collect averages $(1/J)\sum_{k=1}^J \vartheta_k$ and use them to numerically approximate the posterior mean or quantiles – see Figure 1.1 below for an illustration.

We will turn to the study of performance guarantees for MCMC methods later, but for now let us give a concept proof for how to setup Markov chains that have prescribed posterior distributions $\Pi(\cdot|D_N)$ as invariant measures. Recall that the distribution of a Markov chain is characterised by its initial condition ϑ_0 and by its transition probabilities $\Pr(\vartheta_m \in B|\vartheta_{m-1} = t), t \in \Theta, m \in \mathbb{N}$, where B is any (measurable) subset of Θ. By the Markov property the preceding transition probabilities are the same for all m, with Markov kernel $K(t, \cdot)$

$$\Pr(\vartheta_1 \in B|\vartheta_0 = t) = K(t, B), \quad t \in \Theta, B \subset \Theta \text{ (measurable)}. \tag{1.16}$$

A probability measure μ on Θ is called *invariant* (or stationary) for the Markov kernel K if

$$\int_\Theta K(t, B)d\mu(t) = \mu(B) \ \ \forall B \subset \Theta \text{ (measurable)}.$$

In numerical practice we will work with a high-dimensional Euclidean 'approximation' space $\mathbb{R}^D$ of Θ, and wish to generate samples $\vartheta_1, \vartheta_2, \dots, \vartheta_M, \dots$ from a Markov chain $\{\vartheta_m : m \in \mathbb{N}\}$ taking values in $\Theta = \mathbb{R}^D$ that has as invariant measure $\mu = \Pi(\cdot|D_N)$ the posterior distribution.

The pCN algorithm as a Metropolis Hastings algorithm

A general class of so-called Metropolis Hastings algorithms arises in settings where we can evaluate ratios $\mu(s)/q(s|t)$ for some auxiliary conditional probability density function $q(\cdot|t)$ on Θ. In this case we can generate a Markov chain $\{\vartheta_m : m \in \mathbb{N}\}$ as follows:

- 1. For $m \in \mathbb{N}$ and given ϑ_m, generate a new draw (the 'proposal') $s_m \sim q(\cdot|\vartheta_m)$.
- 2. Define

$$\vartheta_{m+1} = \begin{cases} s_m, & \text{with probability } \rho(\vartheta_m, s_m) \\ \vartheta_m, & \text{with probability } 1 - \rho(\vartheta_m, s_m), \end{cases}$$

with 'acceptance' probability

$$\rho(t, s) = \min\left\{\frac{\mu(s)}{\mu(t)}\frac{q(t|s)}{q(s|t)}, 1\right\}.$$

The following proposition is standard and left as Ex. 1.4.2.

Proposition 1.2.1. *Let the Markov chain $\{\vartheta_m : m \in \mathbb{N}\}$ be generated as above and suppose $\mu, q(\cdot|t), t \in \Theta$, are strictly positive throughout Θ. Then μ is an invariant measure for the Markov chain.*

When the prior Π giving rise to (1.14) is a Gaussian $N(0, \Sigma_D)$-distribution on $\Theta = \mathbb{R}^D$ with non-singular covariance matrix Σ_D, we can use it to construct the 'proposal distribution' q. This leads to the following 'pCN' Markov chain:

- 1. For $m \in \mathbb{N}$ and given ϑ_m, generate a random vector $\xi \sim N(0, \Sigma_D)$ and set

$$s_m = \sqrt{1-2\delta}\vartheta_m + \sqrt{2\delta}\xi$$

for some fixed 'step size' $\delta > 0$.
- 2. Define

$$\vartheta_{m+1} = \begin{cases} s_m, & \text{with probability } \rho(\vartheta_m, s_m) \\ \vartheta_m, & \text{with probability } 1 - \rho(\vartheta_m, s_m), \end{cases}$$

with 'acceptance probabilities'

$$\rho(\vartheta_m, s_m) = \min\left\{e^{\ell_N(s_m) - \ell_N(\vartheta_m)}, 1\right\}.$$

An implementation of pCN for the non-linear inverse problem from (1.1) can be found in Figure 1.1 below – this is taken from [91], where further details about the precise statistical measurement model, step size and choice of prior parameters can be found.

The statistical intuition behind this algorithm is simple: We compute the new (conditionally Gaussian) proposal s_m and perform a likelihood ratio test against the previous position ϑ_m of the Markov chain. If the likelihood is increased we always accept the proposal. If the likelihood is decreased we still accept with a certain probability to ensure the Markov chain does *not* stick to a local optimum. The data D_N only enters the algorithm when computing the likelihood ratios. This method targets the correct posterior distribution:

Proposition 1.2.2. *The preceding pCN Markov chain (ϑ_m) has the posterior distribution (1.14) as invariant measure whenever the prior Π follows a $N(0, \Sigma_D)$ distribution on $\mathbb{R}^D = \Theta$.*

Proof. Note that for our choice of prior the target distribution (1.14) is of the form

$$\mu(\theta) \propto \exp\left\{\ell_N(\theta) - \frac{\theta^T \Sigma_D^{-1} \theta}{2}\right\}, \quad \theta \in \mathbb{R}^D, \tag{1.17}$$

where ℓ_N is as in (1.15). We will show that the pCN chain is a special case of a Metropolis Hasting algorithm with proposal density

$$q(\cdot|t) \sim N(\sqrt{1-2\delta}t, 2\delta\Sigma_D) \text{ on } \Theta = \mathbb{R}^D, \quad t \in \Theta.$$

To simplify the algebra we set $\Sigma_D = I_D$ so that the proposal densities are

$$q(s|t) \propto \exp\left\{-\frac{1}{4\delta}\|s - \sqrt{1-2\delta}t\|^2\right\}, \quad s, t \in \mathbb{R}^D \tag{1.18}$$

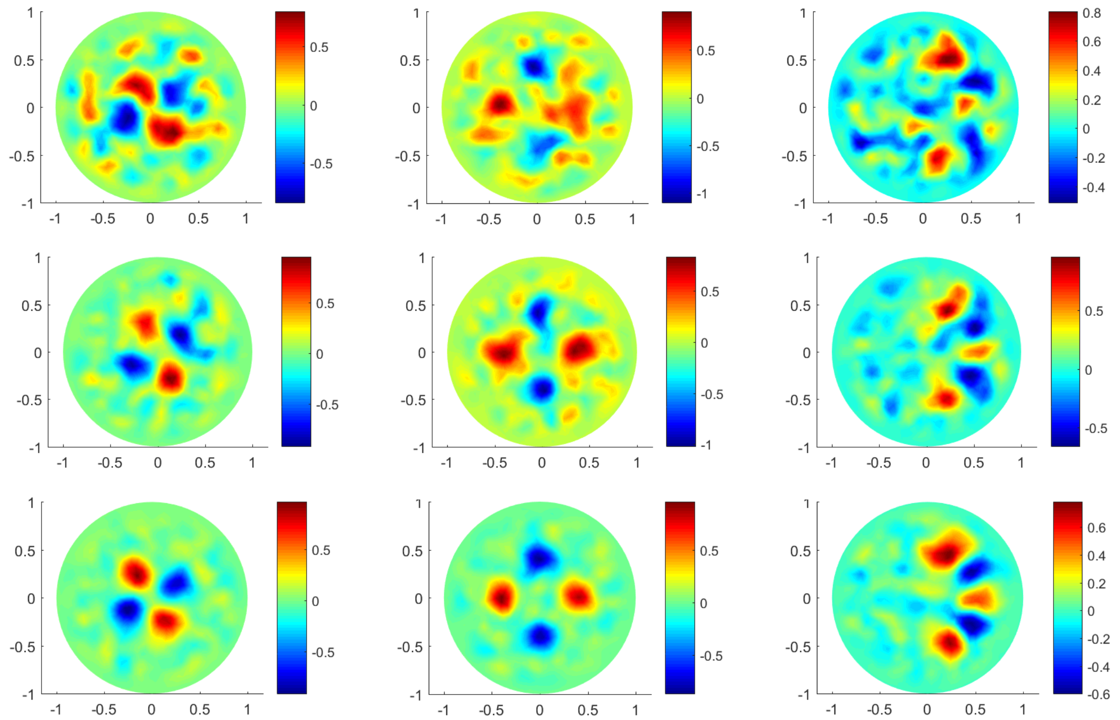
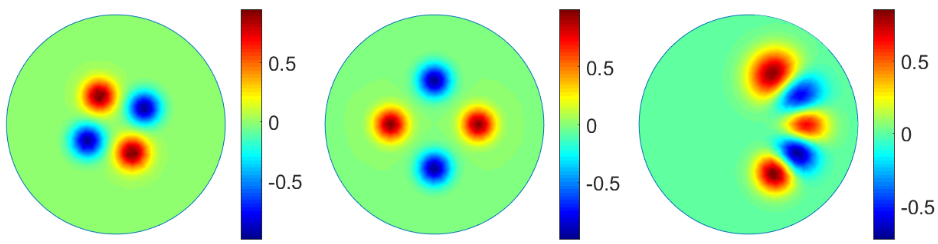


Figure 1.1: Top 3 rows: The 3-dimensional posterior mean field $(1/J)\sum_{k=1}^{J}\vartheta_k$ for sample sizes $N = 200, 400, 800$, arising from a ($so(3)$-valued) Matern Gaussian process prior, applied to a non-Abelian X-ray transform (1.1). The number of MCMC iterations via pCN is $J = 100000$. Bottom: the true matrix field θ_0.



with $\|\cdot\|$ the Euclidean norm on $\mathbb{R}^D$ – but the proof works for any non-singular Σ_D just as well. Conditional sampling from $q(\cdot|t)$ amounts simply to drawing a normal random vector. Basic calculations give

$$\begin{aligned}\frac{\mu(s)}{\mu(t)}\frac{q(t|s)}{q(s|t)} &= \exp\left\{\ell_N(s) - \frac{\|s\|^2}{2} - \ell_N(t) + \frac{\|t\|^2}{2}\right\} \\ &\quad \times \exp\left\{-\frac{1}{4\delta}\|t-\sqrt{1-2\delta}s\|^2 + \frac{1}{4\delta}\|s-\sqrt{1-2\delta}t\|^2\right\} \\ &= e^{\ell_N(s)-\ell_N(t)},\end{aligned}$$

showing that the pCN scheme indeed has the appropriate 'Metropolis-Hastings' correction, and Proposition 1.2.1 applies. □

We conclude that for Gaussian priors, approximate posterior computation by MCMC is in principle possible so long as we can evaluate $\ell_N(\theta)$ at each iteration.

Langevin algorithm

Another approach to sample from a prescribed target measure μ is based on ideas from stochastic differential equations. Here μ does not need to arise from a Gaussian prior as in Proposition 1.2.2 but rather $\mu = \mu_U$ is a general Borel probability measure on $\Theta = \mathbb{R}^D$ which has a Lebesgue density proportional to e^{-U} for some potential $U : \mathbb{R}^D \to \mathbb{R}$, specifically

$$\mu_U(B) = \frac{\int_B e^{-U(\theta)}d\theta}{\int_{\mathbb{R}^D} e^{-U(\theta)}d\theta}, \quad B \subseteq \mathbb{R}^D \text{ measurable.} \tag{1.19}$$

If the gradient ∇U of U is Lipschitz continuous we define the continuous time *Langevin diffusion process* as the unique strong solution $(L_t : t \geq 0)$ of the stochastic differential equation (SDE)

$$dL_t = -\nabla U(L_t)dt + \sqrt{2}dW_t, \quad t \geq 0,\ L_t \in \mathbb{R}^D, \tag{1.20}$$

where $(W_t : t \geq 0)$ is a D-dimensional standard Brownian motion.

Proposition 1.2.3. *The above SDE has a path-wise solution $(L_t : t \geq 0)$ which is a continuous time Markov process with invariant measure μ_U from (1.19).*

Proof. This is a standard result in stochastic calculus. The generator of this Markov process equals the second order elliptic operator $\mathscr{L}_U = \Delta + \nabla U \cdot \nabla$. Then μ_U solves the PDE $\mathscr{L}_B^* \mu_U = 0$ where $\mathscr{L}_B^*$ is the adjoint of $\mathscr{L}_B$ for dx, and hence is an invariant measure. See p.46f. in [7] or also [13] for details. □

The Euler-Maruyama discretisation of the dynamics (1.20) gives rise to the discrete-time Markov chain $(\vartheta_k : k \geq 0)$,

$$\vartheta_{k+1} = \vartheta_k - \gamma \nabla U(\vartheta_k) + \sqrt{2\gamma}\xi_{k+1}, \quad k \geq 0, \tag{1.21}$$

where $(\xi_k : k \geq 1)$ form an i.i.d. sequence of D-dimensional standard Gaussian $N(0, I_{D\times D})$ vectors, $\gamma > 0$ is some fixed *step size*, and ϑ_0 an initial value. By choosing for $U = \ell_N + \log d\Pi$ the appropriate potential from (1.14) we can thus target the posterior measure by the Markov chain $(\vartheta_k : k \in \mathbb{N})$.

Just as in the pCN scheme, running this Markov chain requires drawing a Gaussian variable at each step as well as an evaluation of the gradient of ℓ_N (and hence of $\mathscr{G}, \nabla\mathscr{G}$), but neither the computation of $\mathscr{G}^{-1}$ nor of the normalising factor in (1.14). Note that in contrast to pCN the above algorithm does not include an 'accept/reject' Metropolis-Hastings correction step – the invariant measure of the discrete scheme is thus slightly different from μ_U even though the misspecification error decreases with the discretisation error/step size γ, so that the method is still practical in numerical computation. We will refer to (ϑ_k) as the *unadjusted Langevin algorithm* (ULA) in what follows, and its performance will be studied in detail in Chapter 5.

1.3 The frequentist perspective on posterior inference

While we rely on the Bayesian formalism to construct the posterior distribution $\Pi(\cdot|D_N)$ from which we generate statistical inference algorithms, to obtain performance guarantees we will instead take an 'objective' point of view and study the behaviour of $\Pi(\cdot|D_N)$ under the 'frequentist hypothesis' that the $X_1, X_2, \dots,$ are drawn from the infinite product measure $P_{\theta_0}^{\mathbb{N}}$, where $\theta_0 \in \Theta$ is the 'ground truth parameter'. This view-point dates back to Laplace [79] and is well-studied in mathematical statistics [59, 61, 83, 127] – see also the notes to this section for more on the history of this problem. It also appears most reasonable in the scientific applications of PDEs and inverse problems we have in mind where the prior should be regarded as a regularisation tool rather than a subjective belief. In this section we give a first fundamental result about the frequentist behaviour of posterior measures in the setting of the regression model (1.11). In the proofs we will require some information theoretic inequalities which we derive first.

1.3.1 Information distances for random design regression

Various features of statistical models can be encoded in so called *information distances* on the laws $\{P_\theta : \theta \in \Theta\}$ indexing the statistical 'experiment'. In our model (1.11) these laws are equivalently represented by the probability densities p_θ from (1.12). The first and most classical information distance is the *'Kullback-Leibler (KL)' divergence, or entropy*, given by

$$KL(P_\theta, P_\vartheta) := E_\theta\Big[\log \frac{p_\theta}{p_\vartheta}(Y,X)\Big], \tag{1.22}$$

a quantity that is always non-negative by an application of Jensen's inequality. Often we need to control higher (say, second) moments of the log-ratio in the last expectation, so we introduce

$$V(P_\theta, P_\vartheta) := E_\theta\Big[\log \frac{p_\theta}{p_\vartheta}(Y,X)\Big]^2. \tag{1.23}$$

The *Hellinger distance* h is given by

$$h^2(p_\theta, p_\vartheta) := \int_{V\times\mathcal{X}} \big[\sqrt{p_\theta} - \sqrt{p_\vartheta}\big]^2 d\mu \tag{1.24}$$

The following proposition relates these distances to more analytical aspects of the 'forward' maps $\mathscr{G}(\theta)$ appearing in our regression model (1.11), assuming the latter are uniformly bounded by a fixed constant U.

Proposition 1.3.1. *Suppose that for a subset $\Theta \subset L^2(\mathcal{Z}, W)$ and some finite constant $U = U_{\mathscr{G},\Theta,V} > 0$ we have*

$$\sup_{\theta\in\Theta} \big\| |\mathscr{G}(\theta)(\cdot)|_V \big\|_\infty \le U. \tag{1.25}$$

For the model densities from (1.12) arising from (1.11) we then have for every $\theta, \vartheta \in \Theta$,

$$KL(P_\theta, P_\vartheta) = \frac{1}{2}\|\mathscr{G}(\theta) - \mathscr{G}(\vartheta)\|^2_{L^2_\lambda(\mathcal{X},V)}, \tag{1.26}$$

$$V(P_\theta, P_\vartheta) \le 2(U^2+1)\|\mathscr{G}(\theta) - \mathscr{G}(\vartheta)\|^2_{L^2_\lambda(\mathcal{X},V)}, \tag{1.27}$$

as well as

$$C_U\|\mathscr{G}(\theta) - \mathscr{G}(\vartheta)\|^2_{L^2_\lambda(\mathcal{X},V)} \le h^2(p_\theta, p_\vartheta) \le \frac{1}{4}\|\mathscr{G}(\theta) - \mathscr{G}(\vartheta)\|^2_{L^2_\lambda(\mathcal{X},V)}. \tag{1.28}$$

where

$$C_U = \frac{1 - e^{-U^2/2}}{2U^2}.$$

Proof. For $(Y,X) \sim P_\theta$ from (1.11) with noise variable $\varepsilon \sim N(0, I_V)$, and using (1.12) we see that

$$\log \frac{p_\theta}{p_\vartheta}(Y,X) = \frac{1}{2}|\mathscr{G}(\theta)(X) - \mathscr{G}(\vartheta)(X)|_V^2 + \langle \varepsilon, \mathscr{G}(\theta)(X) - \mathscr{G}(\vartheta)(X)\rangle_V \qquad (1.29)$$

and the $E_\theta = E_\lambda E_\varepsilon$ expectation of the r.h.s. is precisely (1.26) as $E_\varepsilon \varepsilon = 0$.

Next, squaring the last displayed identity and using $(a+b)^2 \le 2(a^2+b^2)$,

$$\begin{aligned} E_\theta\Big[\log \frac{p_\theta}{p_\vartheta}(Y,X)\Big]^2 &\le \frac{1}{2}E_\lambda\Big[|\mathscr{G}(\theta)(X) - \mathscr{G}(\vartheta)(X)|_V^2\Big]^2 + 2E_\lambda E_\varepsilon \langle \varepsilon, \mathscr{G}(\theta)(X) - \mathscr{G}(\vartheta)(X)\rangle_V^2 \\ &\le 2(U^2+1)\|\mathscr{G}(\theta) - \mathscr{G}(\vartheta)\|^2_{L^2_\lambda(\mathcal{X},V)} \end{aligned}$$

where we have used (1.25) and that conditionally on X, the random variable $\langle \varepsilon, \mathscr{G}(\theta)(X) - \mathscr{G}(\vartheta)(X)\rangle_V$ is $N(0, |\mathscr{G}(\theta)(X) - \mathscr{G}(\vartheta)(X)|_V^2)$ distributed. This proves (1.27).

It remains to prove (1.28). We first bound the so called *Hellinger affinity* defined as

$$\rho(p_\theta, p_\vartheta) := \int_{V\times\mathcal{X}} \sqrt{p_\theta p_\vartheta}\, d\mu = 1 - \frac{1}{2}h^2(p_\theta, p_\vartheta), \qquad (1.30)$$

where the second identity is easily verified. By (1.12), writing $\phi_V = (2\pi)^{-p_V/2}e^{-|\cdot|_V^2/2}$ for the (Lebesgue-) probability density of a standard multivariate normal variable Z on the finite-dimensional vector space V, and using the standard identity $Ee^{\langle u, Z\rangle_V} = e^{|u|_V^2/2}, u \in V$, for Laplace transforms of such variables,

$$\begin{aligned} &\rho(p_\theta, p_\vartheta) \\ &= \frac{1}{(2\pi)^{p_V/2}} \int_{V\times\mathcal{X}} \exp\Big\{-\frac{1}{4}|y - \mathscr{G}(\theta)(x)|_V^2 - \frac{1}{4}|y - \mathscr{G}(\vartheta)(x)|_V^2\Big\} d\mu(y,x) \\ &= \int_{\mathcal{X}} \exp\Big\{-\frac{1}{4}(|\mathscr{G}(\theta)(x)|_V^2 + |\mathscr{G}(\vartheta)(x)|_V^2)\Big\} \int_V e^{\frac{1}{2}\langle y, \mathscr{G}(\theta)(x) + \mathscr{G}(\vartheta)(x)\rangle_V} \phi_V(y)\, dy\, d\lambda(x) \\ &= \int_{\mathcal{X}} \exp\Big\{-\frac{2}{8}(|\mathscr{G}(\theta)(x)|_V^2 + |\mathscr{G}(\vartheta)(x)|_V^2) + \frac{1}{8}|\mathscr{G}(\theta)(x) + \mathscr{G}(\vartheta)(x)|_V^2\Big\} d\lambda(x) \\ &= E_\lambda \exp\Big\{-\frac{1}{8}|\mathscr{G}(\theta)(X) - \mathscr{G}(\vartheta)(X)|_V^2\Big\}. \end{aligned}$$

By Jensen's inequality the r.h.s. is lower bounded by

$$\exp\Big\{-\frac{1}{8}\|\mathscr{G}(\theta) - \mathscr{G}(\vartheta)\|^2_{L^2_\lambda(\mathcal{X},V)}\Big\}.$$

Combined with (1.30) and the standard bound $1 - e^{-z} \le z$ for $z \ge 0$ this gives

$$h^2(p_\theta, p_\vartheta)/2 \le 1 - \exp\Big\{-\frac{1}{8}\|\mathscr{G}(\theta) - \mathscr{G}(\vartheta)\|^2_{L^2_\lambda(\mathcal{X},V)}\Big\} \le \frac{1}{8}\|\mathscr{G}(\theta) - \mathscr{G}(\vartheta)\|^2_{L^2_\lambda(\mathcal{X},V)}$$

so that the second inequality in (1.28) follows. For the first inequality we use the basic consequence of convexity of e^{-z},

$$e^{-z_1} \le \frac{e^{-z_2}-1}{z_2} z_1 + 1, \quad 0 \le z_1 < z_2,$$

with choices

$$z_1 = \frac{1}{8}|\mathscr{G}(\theta)(X) - \mathscr{G}(\vartheta)(X)|_V^2, \quad z_2 = U^2/2,$$

permitted in view of (1.25), to the effect

$$E_\lambda \exp\left\{-\frac{1}{8}|\mathscr{G}(\theta)(X) - \mathscr{G}(\vartheta)(X)|_V^2\right\} \le \frac{e^{-U^2/2}-1}{4U^2}\|\mathscr{G}(\theta) - \mathscr{G}(\vartheta)\|^2_{L^2_\lambda(\mathcal{X},V)} + 1.$$

Combined with the displayed identity for $\rho(p_\theta, p_\vartheta)$ and (1.30) this implies

$$1 - \frac{1}{2}h^2(p_\theta, p_\vartheta) \le \frac{e^{-U^2/2}-1}{4U^2}\|\mathscr{G}(\theta) - \mathscr{G}(\vartheta)\|^2_{L^2_\lambda(\mathcal{X},V)} + 1$$

from which the l.h.s. of (1.28) follows immediately. This completes the proof. □

We see from this proposition that relevant information distances are closely related to the standard $L^2_\lambda(\mathcal{X}, V)$-norms, as long as the $\mathscr{G}(\theta)$ are uniformly bounded (without the latter assumption the story can be quite different, see [20]). In the PDE models relevant here the uniform boundedness hypothesis will be satisfied.

1.3.2 A first posterior contraction theorem

Let us introduce the (semi-) metric

$$d_{\mathscr{G}}(\theta, \theta') := \|\mathscr{G}(\theta) - \mathscr{G}(\theta')\|_{L^2_\lambda(\mathcal{X},V)}$$

on Θ. For any subset $\Theta' \subset \Theta$ we denote by $N(\Theta', d_{\mathscr{G}}, \eta)$, $\eta > 0$, the minimal number of balls of $d_{\mathscr{G}}$-radius η needed to cover Θ'.

The proof of the next theorem is based on a general principle that ensures that posterior measures in i.i.d. models contract about Hellinger neighbourhoods of the ground truth $\theta_0 \in \Theta$ that generated the data, as long as the prior is not too rough (i.e., concentrates on sets of not too large metric capacity) and charges a 'neighbourhood' of θ_0 for the KL- and V-distance of the model with sufficiently high probability. By Proposition 1.3.1 this translates into a comparable result in the $d_{\mathscr{G}}$-distance when data arises from (1.11) with uniformly bounded regression functions $\mathscr{G}(\theta)$.

Theorem 1.3.2. *Let Π_N be a sequence of prior Borel probability measures on some Borel subset $\Theta \subset L^2_\zeta(\mathcal{Z}, W)$, and let $\Pi_N(\cdot|(Y_i, X_i)_{i=1}^N) = \Pi_N(\cdot|D_N)$ be the resulting posterior distribution (1.14) arising from observations in model (1.11) with forward map $\mathscr{G} : \Theta \to L^2_\lambda(\mathcal{X}, V)$. Assume that for some fixed $\theta_0 \in \Theta$, envelope constants $U = U_N \geq 1, \|\mathscr{G}(\theta_0)\|_\infty \leq U$, and sequence $\delta_N \to 0$ s.t. $N\delta_N^2 \geq 1$, $\sqrt{N}\delta_N/U_N \to \infty$ as $N \to \infty$, the sets*

$$\mathcal{B}_N := \Big\{\theta \in \Theta : d_{\mathscr{G}}(\theta, \theta_0) \leq \delta_N, \quad \|\mathscr{G}(\theta)\|_\infty \leq U\Big\}, \tag{1.31}$$

satisfy for all N large enough

$$\Pi_N(\mathcal{B}_N) \geq e^{-AN\delta_N^2}, \quad \text{some } A > 0. \tag{1.32}$$

Further assume that there exists a sequence of Borel sets $\Theta_N \subset \Theta$ for which

$$\Pi_N(\Theta_N^c) \leq e^{-BN\delta_N^2}, \quad \text{some } B > A + 2, \tag{1.33}$$

and such that for all $\bar{m} > 0$ large enough

$$\log N(\Theta_N, d_{\mathscr{G}}, \bar{m}\delta_N) \leq N\delta_N^2. \tag{1.34}$$

Then, for all $0 < b < B - A - 2$ we can choose $L = L(B, \bar{m}, b)$ large enough s.t.

$$P_{\theta_0}^N\Big(\Pi_N\big(\theta \in \Theta_N, d_{\mathscr{G}}(\theta, \theta_0) \leq L\delta_N C_{V_N}^{-1}|D_N\big) \leq 1 - e^{-bN\delta_N^2}\Big) \to 0 \tag{1.35}$$

as $N \to \infty$, where the constant C_V is as after (1.28) with $V = V_N = \max(U, \sup_{\theta\in\Theta_N} \|\mathscr{G}(\theta)\|_\infty)$.

Proof. We first need a lemma that we shall also use in different contexts later, and which also clarifies that the denominator in the formula for the posterior measure (1.14) is bounded away from zero on events of high $P_{\theta_0}^{\mathbb{N}}$-probability, so that the posterior ratio is well-defined not just Q-almost surely but also in the frequentist sense.

Lemma 1.3.3. *Let $\mathscr{G}$ be as in the theorem and let ν be a probability measure on some (measurable) subset $B_N \subseteq \mathcal{B}_N$. Then for ℓ_N from (1.15) and all $K \geq 2$ we have*

$$P_{\theta_0}^N\left(\int_{B_N} e^{\ell_N(\theta)-\ell_N(\theta_0)} d\nu(\theta) \leq e^{-KN\delta_N^2}\right) \leq \frac{8(U^2+1)}{K^2N\delta_N^2}. \tag{1.36}$$

Proof. Note first that by (1.27), $\log(p_\theta/p_{\theta_0})(Y, X)$ is square-integrable for the product measure $P_{\theta_0} \otimes \nu$. Then from Jensen's inequality (applied to log and $\int(\cdot)d\nu$) and recalling (1.12), the probability in question is bounded by

$$P_{\theta_0}^N\left(\int\int_{B_N} \log\frac{p_\theta}{p_{\theta_0}} d\nu(\theta) d(P_N - P_{\theta_0}) \leq -K\delta_N^2 - \int\int_{B_N} \log\frac{p_\theta}{p_{\theta_0}} d\nu(\theta) dP_{\theta_0}\right)$$

where $P_N = (1/N)\sum_{i=1}^N \delta_{(Y_i X_i)}$ is the empirical measure associated to the sample from P_{θ_0}. Now as in (1.29) and by definition of $\mathcal{B}_N$, for all $\theta \in B_N$,

$$-\int \log \frac{p_\theta}{p_{\theta_0}} dP_{\theta_0} = \frac{1}{2}\|\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\|^2_{L^2_\lambda(\mathcal{X},V)} \le \delta_N^2$$

so that for $K \ge 2$ and using also Fubini's theorem, the last probability can be bounded by

$$P^N_{\theta_0}\left(\sqrt{N}\int\int_{B_N} \log \frac{p_{\theta_0}}{p_\theta} d\nu(\theta) d(P_N - P_{\theta_0}) \ge K\sqrt{N}\delta_N^2/2\right).$$

Now using the inequalities of Chebyshev and Jensen, Fubini's theorem, (1.27) from Proposition 1.3.1 and again the definition of $\mathcal{B}_N \supset B_N$, the last probability is bounded as

$$\frac{4E_{\theta_0}\big[\int_{B_N}(\log(p_{\theta_0}/p_\theta)(Y,X)d\nu(\theta)\big]^2}{K^2 N\delta_N^4} \le \frac{4\int_{B_N} V(P_{\theta_0}, P_\theta)d\nu(\theta)}{K^2 N \delta_N^4} \le \frac{8(U^2+1)}{K^2 N\delta_N^2},$$

completing the proof of the lemma. $\square$

We now prove the theorem. In view of the last lemma with $\nu = \Pi(\cdot)/\Pi(B_N), B_N = \mathcal{B}_N, K = A+2$, and (1.32) we can restrict to events

$$A_N = \left\{\int_\Theta e^{\ell_N(\theta)-\ell_N(\theta_0)}\, d\Pi(\theta) \ge e^{-(A+2)N\delta_N^2}\right\} \tag{1.37}$$

of probability $P^N_{\theta_0}(A_N) \to 1$ as $N \to \infty$. Moreover using (1.34) and the r.h.s. in (1.28) we can verify the metric entropy condition in [61, Theorem 7.1.4] with choices $\varepsilon_0 = m'\delta_N$ and $\log N(\varepsilon) = cN\delta_N^2, c > 0$, constant in $\varepsilon > \varepsilon_0$, to deduce that for every $k > 1$ we can choose $m' < m$ large enough and find 'tests' $\Psi_N = \Psi : (V \times \mathcal{X})^N \to \{0,1\}$ for which

$$P^N_{\theta_0}(\Psi_N = 1) \to_{N\to\infty} 0 \text{ and } \sup_{\theta\in\Theta_N: h(p_\theta, p_{\theta_0}) > m\delta_N} E^N_\theta(1-\Psi_N) \le e^{-kN\delta_N^2}. \tag{1.38}$$

Now let us write

$$\bar\Theta_N = \Theta_N \cap \{h(p_\theta, p_{\theta_0}) \le m\delta_N\}$$

and $\bar\Theta_N^c$ for its complement in Θ. By (1.14), as $N \to \infty$,

$$\begin{aligned}
&P^N_{\theta_0}\big(\Pi\big(\bar\Theta^c_N|D_N\big) \ge e^{-bN\delta_N^2}\big)\\
&= P^N_{\theta_0}\left(\frac{\int_{\bar\Theta^c_N} e^{\ell_N(\theta)-\ell_N(\theta_0)} d\Pi(\theta)}{\int_\Theta e^{\ell_N(\theta)-\ell_N(\theta_0)} d\Pi(\theta)} \ge e^{-bN\delta_N^2}, \Psi_N = 0, A_N\right) + o(1)\\
&\le P^N_{\theta_0}\left(\int_{\bar\Theta^c_N} e^{\ell_N(\theta)-\ell_N(\theta_0)} d\Pi(\theta)(1-\Psi_N) \ge e^{-(b+A+2)N\delta_N^2}\right) + o(1).
\end{aligned}$$

By Markov's inequality, decomposing

$$\bar{\Theta}_N^c = \Theta_N^c \cup \{h(p_\theta, p_{\theta_0}) > m\delta_N\},$$

and using Fubini's theorem as well as

$$E_{\theta_0}^N \prod_{i=1}^N \frac{p_\theta}{p_{\theta_0}}(1-\Psi_N) = E_\theta^N(1-\Psi_N) \le 1 \tag{1.39}$$

we further bound the last probability as

$$\begin{aligned}
&e^{(b+A+2)N\delta_N^2} \int_{\bar{\Theta}_N^c} E_\theta^N(1-\Psi_N) d\Pi(\theta) \\
&\le e^{(b+A+2)N\delta_N^2} \left(2\Pi(\Theta_N^c) + \int_{\theta \in \Theta_N : h(p_\theta, p_{\theta_0}) > m\delta_N} E_\theta^N (1-\Psi_N) d\Pi(\theta) \right) \\
&\lesssim e^{(b+A+2-B)N\delta_N^2} + e^{(b+A+2-k)N\delta_N^2} \to_{N\to\infty} 0,
\end{aligned}$$

where we have used (1.33) and (1.38) with k and then m large enough. The proof of the theorem is completed after noting that the Hellinger-distance contraction result implies the on in $d_{\mathscr{G}}$-distance by virtue of the first inequality in (1.28) with appropriate envelope constant. □

1.4 Notes

1.4.1 Exercises

Exercise 1.4.1. Show that the solution maps $f \mapsto u_f$ for (1.2) with $\mathscr{L}_f$ given by (1.3), (1.4), respectively, are *not linear*. [Hint: Initially consider the case $d = 1$ and then for $d > 1$ take $f(x_1, \dots, x_d) = f(x_1)$. For the Schrödinger equation this can also be inferred from the Feynman-Kac formula (6.27).]

Exercise 1.4.2. Prove Proposition 1.2.1. [See, e.g., [113].]

1.4.2 Remarks and comments

Inverse problems and parameter identification problems for PDEs have been studied for a long time - see [52, 69, 74, 75, 105, 119] and references therein. The PDE (1.2) with diffusion operator $\mathscr{L}_f$ from (1.3) serves as 'fruitfly' example throughout large parts of the literature (see Section 3.7 in [119] and [63] for many more references). The example arising with Schrödinger operators (1.4) has its own physical motivation (see [8, 9, 76]) but equally importantly provides a template for other 'perturbed' differential operators of the form $\mathscr{D} + f$, where $\mathscr{D}$ is not necessarily elliptic or of second order. In fact the theory laid out in these notes has been demonstrated to be compatible with the analytically more challenging example of *non-Abelian X-ray transforms* from (1.1), see [91], [92], [23] for references. Here one can regard $\mathscr{D}$ to be the geodesic vector field on a simple manifold. Many more such 'geometric' non-linear inverse problems are discussed in detail in [75, 105] and serve as further examples. But the focus of these notes is primarily on *statistical* aspects of the problem and the analytically 'softer' elliptic model examples already exhibit many of the main statistical features of the theory.

Random design regression models (1.11) are convenient in our setting because of the i.i.d. structure of the samples (Y_i, X_i), which will later on permit the use of techniques from empirical process theory and concentration of measure [48, 61, 121, 130, 132]. They also provide a natural link between standard information distances and the L^2-structure on the 'forward data' $\mathscr{G}(\theta)$ – Proposition 1.3.1 is essentially due to [20]. Regarding the X_i as random is a popular approach studied independently in 'probabilistic numerics', see [43] and more recently [25]. Random design regression models can be shown to be 'asymptotically equivalent' to most other commonly used nonparametric regression models, see [110], and our findings thus also inform such measurement settings. We further note that the theory in this manuscript works as well if the noise vector ε_i is not Gaussian but has mean zero $E\varepsilon_i = 0$ and sub-Gaussian moments. As long as we still use the (now mis-specified) 'quasi-likelihood' model arising from Gaussian densities (1.12), one can prove a version of Theorem 1.3.2 following ideas in [59], Section 8.5.2. A similar remark about the noise variables applies to developments in later chapters and will not be repeated.

Key references for the Bayesian approach to inverse problems were discussed at the beginning of this chapter. MCMC methods have been highly influential in statistical science in the last few decades, see [113] for an introductory text and [36], [66], [18] in the context of the infinite-dimensional models relevant here. See specifically [66] for a more detailed theoretical treatment of the pCN algorithm in this context. Langevin MCMC methods also have a long history [114] and recent interest has been triggered in high-dimensional settings by the articles [39, 49] which will be relevant in Chapter 5 below.

Theorem 1.3.2 is a variant of a by now classical contraction rate result for posterior distributions in infinite-dimensional statistical parameter spaces first given in the landmark paper [58]. An in depth account of this theory can be found in [59] and what is relevant for the proofs here can also be found in Chapter 7.3 in [61]. The proofs naturally exploit properties of the Hellinger distance which plays a central role throughout mathematical statistics, see [19, 59, 61, 83, 125].

In traditional 'direct' models, such posterior 'consistency' theorems have a long history with important contributions by Doob, Le Cam, Schwartz, Freedman, Diaconis, Barron, Shervish, Wasserman, Ghosal, van der Vaart among others, see [46], [82], [55], [117], [44], [12], [58] and also Section 6.9 in [59] for a precise historical discussion.

Chapter 2

Global stability and posterior consistency

In this section we investigate when the Bayesian posterior distribution solves a non-linear problem successfully in the sense that it will converge in the large sample limit ($N \to \infty$) to a Dirac measure at the 'ground truth' parameter θ_0 generating the data – a notion called *posterior consistency*. We will quantify that rate of convergence in norms on the parameter space Θ of interest. Just as in Theorem 1.3.2, this will involve probabilistic statements under the law $P_{\theta_0}^{\mathbb{N}}$.

The main challenge compared to classical 'direct' statistical problems here is the presence of the *non-linear* map $\mathscr{G}$: Firstly, Bayesian inference can be expected to recover $\mathscr{G}(\theta)$ if an appropriate prior for the regression function $\mathscr{G}(\theta)$ is used (verifying (1.32)). In our setting one will initially employ a – say, Gaussian – prior for the parameter θ of interest. It is not clear whether the implied prior for the regression functions $\mathscr{G}(\theta)$ is still adequate, for instance the prior for $\mathscr{G}(\theta)$ is not Gaussian any longer when $\mathscr{G}$ is non-linear. This issue will be addressed by a 'forward regularity' condition which basically requires Lipschitz continuity of $\mathscr{G}$ for correct norms on a sufficiently large portion Θ' of the parameter space Θ, and which can be verified for the examples considered here by regularity estimates for inhomogeneous elliptic PDEs, as we will show.

The second challenge goes more directly to the heart of the subject: If the 'forward map' $\mathscr{G}$ satisfies what is commonly called a *stability estimate*

$$\|\theta_1 - \theta_2\| \leq c(\Theta')\|\mathscr{G}(\theta_1) - \mathscr{G}(\theta_2)\|^{\eta}, \quad \theta_1, \theta_2 \in \Theta', \tag{2.1}$$

for suitable norms, sufficiently large subsets $\Theta' \subset \Theta$ and 'stability' Hölder exponent $\eta > 0$, then there may be hope that a Bayes method that successfully infers the regression maps $\mathscr{G}(\theta)$ also recovers the parameter θ of interest in the inverse problem. Such stability estimates can be regarded as 'quantitative statements' about the injectivity of the map $\theta \mapsto \mathscr{G}(\theta)$ and hence require a deeper understanding of the analytical properties of the inverse problem at hand. They imply in particular statistical *identifiability* of the model $\{P_\theta : \theta \in \Theta\}$, but for non-linear $\mathscr{G}$ the implied constant $c(\Theta')$ typically grows with the diameter of Θ' and the stability estimate does not hold on the entire parameter space Θ. This causes difficulties in the application of techniques from Bayesian non-parametric statistics with the unbounded parameter spaces that naturally support Gaussian priors.

We will show how the preceding programme can be made to work nevertheless for a class of Gaussian process priors (infinite-dimensional normal distributions). We start with the analytical properties that the forward map $\mathscr{G}$ will be required to verify and check them for our main PDE examples. We then draw from the theory of Gaussian measures in infinite-dimensions to apply Theorem 1.3.2 with appropriate 'regularisation sets' Θ_N, and ultimately deduce a variety of posterior consistency theorems.

2.1 Analytical hypotheses on the forward map and PDE examples

Recall the forward map $\mathscr{G} : \Theta \to L^2_\lambda(\mathcal{X}, V)$ from (1.10), where the domain Θ is assumed to be a (measurable) subset of $L^2_\zeta(\mathcal{Z}, W)$. We introduce a norm $\|\cdot\|_{\mathcal{R}}$ of some linear *regularisation subspace* $\mathcal{R}$ *of* $L^2_\zeta(\mathcal{Z}, W)$ whose choice will later be related to the prior Π used. One could think of $\mathcal{R}$ as a Sobolev or Hölder space of functions over a domain $\mathcal{Z}$. We write

$$B_{\mathcal{R}}(M) = \{\theta \in \mathcal{R} : \|\theta\|_{\mathcal{R}} \le M\} \tag{2.2}$$

for the ball of radius M in the space $\mathcal{R}$.

2.1.1 Forward regularity conditions for $\mathscr{G}$

For $\mathcal{Z}$ a bounded smooth domain in $\mathbb{R}^d$ the standard Sobolev spaces $H^\kappa(\mathcal{Z})$ of W-valued vector fields are naturally defined, and so are their topological dual spaces $(H^\kappa(\mathcal{Z}))^*$ – see Section 6.1.1 below. When $\kappa = 0$ we have $H^0(\mathcal{Z}) = L^2_\zeta(\mathcal{Z}, W)$ and note that then $\|\cdot\|_{(H^0(\mathcal{Z}))^*} = \|\cdot\|_{L^2_\zeta}$ in (2.4) by self-duality of Hilbert spaces.

Condition 2.1.1. *Consider a parameter space $\Theta \subseteq L^2_\zeta(\mathcal{Z}, W)$ and measurable map $\mathscr{G} : \Theta \to L^2_\lambda(\mathcal{X}, V)$. Let $\kappa \ge 0$. Suppose for all $M > 0$ and some normed linear subspace $(\mathcal{R}, \|\cdot\|_{\mathcal{R}})$ of $L^2_\zeta(\mathcal{Z}, W)$, there exist finite constants $U \ge 1$ and $L > 0$ (that may depend on M) such that*

$$\sup_{\theta \in \Theta \cap B_{\mathcal{R}(M)}} \sup_{x \in \mathcal{X}} |\mathscr{G}(\theta)(x)|_V \le U, \ \textit{and} \tag{2.3}$$

$$\|\mathscr{G}(\theta_1) - \mathscr{G}(\theta_2)\|_{L^2_\lambda(\mathcal{X},V)} \leq L\|\theta_1 - \theta_2\|_{(H^\kappa(\mathcal{Z}))^*} \quad \forall \theta_1, \theta_2 \in \Theta \cap B_{\mathcal{R}}(M). \tag{2.4}$$

The condition requires that on bounded subsets of $\mathcal{R}$, the forward map $\mathscr{G}$ is uniformly bounded and Lipschitz from $(H^\kappa)^*$ to L^2_λ. As we are taking a dual space, for $\kappa > 0$ the dual norm is weaker than the standard $L^2_\zeta(\mathcal{Z}, W)$-norm on Θ (when ζ is Lebesgue measure on $\mathcal{Z}$). One can regard κ as measuring the 'forward smoothing' nature of $\mathscr{G}$. Determining the correct value of κ can be challenging in non-linear inverse problems but the results that follow can be used for any choice of $\kappa \geq 0$. [When $\kappa = 0$ we can slightly weaken the requirement (2.4), see Exercise 2.4.4.] The theory also easily generalises to $\mathscr{G}$ that is only Hölder instead of Lipschitz continuous, but this will not be relevant in these notes.

Condition 2.1.1 and PDE examples

To use Gaussian process priors we will require Θ in Condition 2.1.1 to be a *linear* space. For the model PDE from (1.2), the parameter f indexing the differential operators $\mathscr{L}_f$ in (1.3) and (1.4) does not lie in a linear space (as $f \geq f_{min}$ or $f \geq 0$ are required), but we can map a linear space of θ's into positive f's by use of a 'link' function (such as $f = f_\theta = e^\theta > 0$). The forward map $\mathscr{G}$ is then the solution map of these PDEs composed with the link function.

As for these PDEs both parameters θ, f_θ and solutions $\mathscr{G}(\theta)$ are all real-valued functions defined on the same bounded smooth domain $\mathcal{X} \subset \mathbb{R}^d$, we will take $\mathcal{X} = \mathcal{Z}$, $V = W = \mathbb{R}$, and $\lambda = \zeta$ equal to Lebesgue measure normalised to $\lambda(\mathcal{X}) = 1$ when verifying Condition 2.1.1 in what follows. We will use standard techniques from the theory of elliptic PDEs reviewed in Section 6.1 below.

Schrödinger equation

Let us first show how this works in the elliptic PDE (1.2) where

$$\mathscr{L}_f = \frac{\Delta}{2} - f$$

is a Schrödinger operator (1.4). We choose a parameter space $\Theta \subset H^\xi(\mathcal{X}) \subset C(\mathcal{X})$ for some $\xi > d/2$, and enforce positivity of the potential $f \in H^\xi$ by re-parameterising $f_\theta = e^\theta > 0$. The forward map is given by the unique solution

$$\mathscr{G}(\theta) \equiv u_\theta, \quad \theta \in \Theta, \tag{2.5}$$

of (1.2) with $\mathscr{L}_f = \mathscr{L}_{f_\theta}$, smooth boundary 'temperatures' $h \geq h_{min} > 0$ and with vanishing source $g = 0$. The Feynman-Kac formula (6.27) (for standard Brownian motion $\gamma = 1/2$ and $V = f_\theta \geq 0$) implies (cf. (6.30))

$$\sup_{\theta\in\Theta} \|\mathscr{G}(\theta)\|_\infty = \sup_{\theta\in\Theta} \|u_\theta\|_\infty \leq \|h\|_\infty \equiv U < \infty, \tag{2.6}$$

which verifies (2.3) without even specifying a regularisation space $\mathcal{R}$. To check the Lipschitz property (2.4), note that for any $\theta_i \in C(\mathcal{X})$ we necessarily have $u_{\theta_1} - u_{\theta_2} = h - h = 0$ on $\partial\mathcal{X}$ as well as

$$\mathscr{L}_{f_{\theta_1}}[u_{\theta_1} - u_{\theta_2}] = (\mathscr{L}_{f_{\theta_2}} - \mathscr{L}_{f_{\theta_1}})u_{\theta_2} = (f_{\theta_1} - f_{\theta_2})u_{\theta_2} \quad \text{on } \mathcal{X},$$

in other words $w = \mathscr{G}(\theta_1) - \mathscr{G}(\theta_2)$ solves the inhomogeneous Schrödinger equation $\mathscr{L}_{f_{\theta_1}} w = (f_{\theta_1} - f_{\theta_2})u_{\theta_2}$ on $\mathcal{X}$ subject to Dirichlet boundary conditions $w = 0$ on $\partial\mathcal{X}$. The inverse Schrödinger operator $\mathscr{L}_f^{-1}$ for these boundary conditions is a Lipschitz operator on $L^2_\lambda(\mathcal{X})$ with uniform in $f \geq 0$ Lipschitz-constant $c > 0$ – see (6.12) below with $\gamma = 1/2, V = f$. Hence, using also (2.6),

$$\begin{aligned} \|\mathscr{G}(\theta_1) - \mathscr{G}(\theta_2)\|_{L^2_\lambda} &= \|\mathscr{L}_{f_{\theta_1}}^{-1}[(f_{\theta_1} - f_{\theta_2})u_{\theta_2}]\|_{L^2_\lambda} \\ &\leq c\|u_{\theta_2}\|_\infty \|f_{\theta_1} - f_{\theta_2}\|_{L^2_\lambda} \\ &\leq C(c, M)\|h\|_\infty \|\theta_1 - \theta_2\|_{L^2_\lambda} \end{aligned}$$

where $M \geq \max_i \|\theta_i\|_\infty$ is required only in the last step to control the Lipschitz constant of e^θ. [Choosing a positive 'link' function $\phi : \mathbb{R} \to [0, \infty)$ with $\|\phi'\|_\infty < \infty$ (instead of the exponential map), one can obtain $\mathscr{G}$ that is in fact globally Lipschitz on $\Theta = H^\xi$.] Let us summarise this in the following result.

Proposition 2.1.2. *Let $\Theta \subset H^\xi(\mathcal{X})$ for some $\xi > d/2$. The forward map $\mathscr{G}$ from (2.5) satisfies Condition 2.1.1 for $\kappa = 0$ and for any regularisation norm $\|\cdot\|_{\mathcal{R}}$ that dominates the $\|\cdot\|_\infty$-norm.*

What precedes is already useful and general and will be used in later chapters, even though in terms of forward Lipschitz estimates it is not fully sharp: The forward map $\mathscr{G}$ is actually smoothing of order two and Condition 2.1.1 can be checked for $\kappa = 2$ instead of just $\kappa = 0$ (unless the boundary temperatures h are 'irregular'). In this case, however, the non-linearity of the map becomes more apparent and stronger regularisation norms $\|\cdot\|_{\mathcal{R}}$ are required to control the Lipschitz constants in the preceding estimates. We examine how this works for the diffusion equation and leave the details for the Schrödinger case to the reader – see Execise 2.4.1 of this section.

Diffusion equation

We now study solutions to the equation (1.2) with divergence form operator $\mathscr{L}_f = \nabla \cdot (f\nabla)$ from (1.3). As parameter space we consider a subset of the Sobolev space $\Theta \subset H^\beta(\mathcal{X}), \beta > 1 + d/2$. While weaker hypotheses on β are conceivable we assume sufficient regularity to streamline the exposition. For a fixed positive scalar $f_{min} > 0$ we then parameterise positive conductivities as

$$f_\theta := f_{min} + e^\theta, \quad \theta \in H^\beta(\mathcal{X}), \tag{2.7}$$

ensuring strict ellipticity of the operator $\mathscr{L}_{f_\theta}$. The forward map is then given by the solution

$$\mathscr{G}(\theta) \equiv u_{f_\theta}, \quad \theta \in \Theta, \tag{2.8}$$

of (1.2) with $\mathscr{L}_f = \mathscr{L}_{f_\theta}$ and smooth g, h.

To check Condition 2.1.1, we can infer the uniform boundedness (2.3) from the Feynman-Kac representation (6.27) with $\gamma = f_\theta, V = 0$ (see also (6.30)),

$$\|u_{f_\theta}\|_\infty \leq c(\mathcal{X}, d, f_{min})\|g\|_\infty + \|h\|_\infty \equiv U < \infty. \tag{2.9}$$

For (2.4) notice that for any distinct $\theta_i \in \Theta$ and on $\mathcal{X}$,

$$\begin{aligned} \mathscr{L}_{f_{\theta_1}}[\mathscr{G}(\theta_1) - \mathscr{G}(\theta_2)] &= g - g + (\mathscr{L}_{f_{\theta_2}} - \mathscr{L}_{f_{\theta_1}})[u_{f_{\theta_2}}] \\ &= \nabla \cdot ((f_{\theta_2} - f_{\theta_1})\nabla u_{f_{\theta_2}}), \end{aligned} \tag{2.10}$$

and also that $\mathscr{G}(\theta_1) - \mathscr{G}(\theta_2) = h - h = 0$ on $\partial\mathcal{X}$. Thus $\mathscr{G}(\theta_1) - \mathscr{G}(\theta_2)$ itself solves an equation of the type (1.2) with operator $\mathscr{L}_f = \mathscr{L}_{f_{\theta_1}}$, boundary values $h = 0$ and source g given in the r.h.s. in the last display and can hence be represented via the inverse $\mathscr{L}_{f_{\theta_1}}^{-1}$ of $\mathscr{L}_{f_{\theta_1}}$ for Dirichlet boundary conditions. Using the elliptic regularity estimate from Proposition 6.1.4 (with $\gamma = f_{\theta_1}, V = 0$), we have

$$\begin{aligned} \|\mathscr{G}(\theta_1) - \mathscr{G}(\theta_2)\|_{L^2_\lambda} &= \left\| \mathscr{L}_{f_{\theta_1}}^{-1}[\nabla \cdot ((f_{\theta_2} - f_{\theta_1})\nabla u_{f_{\theta_2}})] \right\|_{L^2_\lambda} \\ &\leq C(M) \left\| \nabla \cdot ((f_{\theta_2} - f_{\theta_1})\nabla u_{f_{\theta_2}}) \right\|_{(H_0^2)^*} \end{aligned} \tag{2.11}$$

so long as the regularisation norm $\|\cdot\|_{\mathcal{R}}$ bounds the C^1-norm of θ to the effect that (using also (2.7)) $\sup_{\theta \in B_{\mathcal{R}}(M)} \|f_\theta\|_{C^1} \leq C(M) < \infty$.

We next notice that Proposition 6.1.5 with $\gamma = f_\theta \in H^\beta$ for $\beta > 1 + d/2$ and $V = 0$ combined with the Sobolev imbedding $H^{\beta+1} \subset C^2$ imply that $\sup_{\theta \in B_{\mathcal{R}}(M)} \|u_{f_\theta}\|_{C^2} < \infty$ whenever the $\|\cdot\|_{\mathcal{R}}$-norm dominates that H^β-norm. Then, applying the divergence theorem (6.7) to the vector field $(f_{\theta_2} - f_{\theta_1})\nabla u_{f_{\theta_2}}$, we have

$$\begin{aligned} \left\|\nabla \cdot ((f_{\theta_2} - f_{\theta_1})\nabla u_{f_{\theta_2}})\right\|_{(H_0^2)^*} &= \sup_{\varphi \in H_0^2,\, \|\varphi\|_{H^2} \leq 1} \left| \int_{\mathcal{X}} \varphi \nabla \cdot ((f_{\theta_2} - f_{\theta_1})\nabla u_{f_{\theta_2}}) \right| \\ &= \sup_{\varphi \in H_0^2,\, \|\varphi\|_{H^2} \leq 1} \left| \int_{\mathcal{X}} (f_{\theta_2} - f_{\theta_1})\nabla\varphi \cdot \nabla u_{f_{\theta_2}} \right| \\ &\leq \|f_{\theta_2} - f_{\theta_1}\|_{(H^1)^*} \sup_{\|\varphi\|_{H^2} \leq 1} \|\nabla\varphi \cdot \nabla u_{f_{\theta_2}}\|_{H^1} \\ &\lesssim \|u_{f_{\theta_2}}\|_{C^2} \|f_{\theta_2} - f_{\theta_1}\|_{(H^1)^*} \end{aligned}$$

using also (6.3). Finally, for $\max_i \|\theta_i\|_{C^1} \le M$ the series $\sum_{k=1}^{\infty}(\theta_1 - \theta_2)^{k-1}/k!$ converges (absolutely) to a function that is uniformly bounded in C^1, and by (6.3) we see

$$\|f_{\theta_2} - f_{\theta_1}\|_{(H^1)^*} = \sup_{\|\varphi\|_{H^1}\le 1} \Big| \int_{\mathcal{X}} (\theta_2 - \theta_1)\varphi e^{\theta_1} \sum_{k=1}^{\infty} \frac{(\theta_2-\theta_1)^{k-1}}{k!} \Big| \lesssim \|\theta_1 - \theta_2\|_{(H^1)^*}.$$

In summary this gives

$$\|\mathscr{G}(\theta_1) - \mathscr{G}(\theta_2)\|_{L^2_\lambda} \lesssim c(M)\|\theta_1 - \theta_2\|_{(H^1)^*}$$

and since the H^β-norm ($\beta > 1 + d/2$) dominates the C^1-norm by (6.4) we have:

Proposition 2.1.3. *Let $\Theta \subset H^\beta(\mathcal{X})$ for some $\beta > 1 + d/2$. The forward map $\mathscr{G}$ from (2.8) satisfies Condition 2.1.1 for $\kappa = 1$ and for any regularisation norm $\|\cdot\|_{\mathcal{R}}$ that dominates the $H^\beta(\mathcal{X})$-norm.*

2.1.2 Injectivity and stability estimates

A statistical model of probability distributions $\{P_\theta : \theta \in \Theta\}$ is called *identifiable* if $P_\theta = P_{\theta'}$ can only occur whenever $\theta = \theta'$. In the case of observations arising from (1.11), with uniformly bounded $\mathscr{G}(\theta)$ as in (2.3), this is equivalent to injectivity of the map $\theta \mapsto \mathscr{G}(\theta)$ on the parameter space Θ (as follows, e.g., from (1.28)). To obtain statistical guarantees for posterior distributions, mere injectivity is not enough and we need to quantify in a certain sense 'how injective' the map $\mathscr{G}$ is.

Consider first the forward map (2.5) arising from solutions u_f to the Schrödinger equation. As a consequence (6.29) of the Feynman-Kac formula there exists $c = c(\mathcal{X}, d)$ such that $u_f \ge h_{min} e^{-c\|f\|_\infty} > 0$, and so we can divide by u_f and represent

$$f = \frac{\Delta u_f}{2u_f} \quad \text{on } \mathcal{X}, \tag{2.12}$$

in particular if $u_f = u_g$ then $f = g$ and the map $f \mapsto u_f$ is injective (from $H^\xi \to H^{\xi+2}$, say, cf. (6.11)). If $\theta_0 = \log f_0, f_0 > 0$, is the true parameter, the last identity suggests a stability estimate of the form

$$\|\theta - \theta_0\|_{L^2} \lesssim \|f_\theta - f_0\|_{L^2} \lesssim \|u_{f_\theta} - u_{f_{\theta_0}}\|_{H^2}, \quad \theta \in \Theta = H^\xi, \ \xi > d/2, \tag{2.13}$$

but the constants in these inequalities are not uniform in $\theta \in H^\xi$, among other reasons because the lower bound for u_{f_θ} deteriorates as $e^{e^{\|\theta\|_\infty}}$ for large $\|\theta\|_\infty$. This illustrates why regularisation is important: if we know θ is bounded in a suitable $\mathcal{R}$-norm, we can control the constants in these inequalities and then interpolate the H^2-norms (using (6.5) below) to bound them by a constant factor times the L^2-norm raised to some Hölder exponent $\eta < 1$. This motivates the following hypothesis on the 'inverse modulus of continuity' of $\mathscr{G}$ as in (2.1).

Condition 2.1.4 (Stability estimate). *Let $B_{\mathcal{R}}(M)$ be as in (2.2) for regularity space $\mathcal{R}$. For some $\eta > 0$ and all $M > 0$ suppose there exists a constant $L' = L'_{\mathscr{G}}$ s.t. for all $\delta > 0$ small enough*

$$\sup\Big\{\|\theta - \theta_0\|_{L^2_\zeta} : \ \theta \in \Theta \cap B_{\mathcal{R}}(M), \|\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\|_{L^2_\lambda} \leq \delta\Big\} \leq L'\delta^\eta. \tag{2.14}$$

For the Schrödinger equation this condition can be checked (see Ex. 2.4.1), using the preceding ideas. Condition 2.1.4 is relevant in non-linear inverse problems more generally as only a 'stability inequality' (2.14) is required rather than an inversion formula for $\mathscr{G}^{-1}$ such as (2.12).

Stability estimate for Darcy's problem

We will now prove a stability estimate for the forward map (2.8) arising with the diffusion operator from (1.3). The inverse problem can be cast as determining a solution $f = f_\theta$ in the first order PDE

$$\nabla f \cdot v + af = g \ \text{ on } \mathcal{X}, \tag{2.15}$$

given $v = \nabla u, a = \Delta u, g$ and boundary values $u = 0$ (say). When ∇u does not vanish (e.g., as in Ex. 2.4.2), we could use the method of characteristics to solve this equation in f along the flow lines (integral curves) of the vector field ∇u (where it becomes an ODE). The hypothesis that ∇u does not vanish is however rather atypical given that u arises as a solution to (1.2), and when ∇u has zeros this approach is not appropriate. While this issue will resurface in later chapters, for now we note that when $g > 0$, then for any $x \in \mathcal{X}$ where $\nabla u(x) = 0$ we must have $\Delta u(x) > 0$, and this can be used to show injectivity of $\mathscr{G}$ by an argument using 'integrating factors'. To accommodate such scenarios in a unified way we will require the hypothesis (2.16), discussed in some more detail after the proof.

Proposition 2.1.5. *Let $\theta_1, \theta_2 \in H^\beta(\mathcal{X}), \beta > 1 + d/2$, be such that $\|\theta_i\|_{C^1} \leq B$, $\theta_1 = \theta_2$ on $\partial\mathcal{X}$, and denote by u_{f_θ} the corresponding solutions to (1.2) with $\mathscr{L}_{f_\theta}$ from (1.3), f_θ as in (2.7), and for smooth g, h. Assume moreover*

$$\inf_{x\in\mathcal{X}} \left[\frac{1}{2}\Delta u_{f_\theta}(x) + \mu|\nabla u_{f_\theta}(x)|^2_{\mathbb{R}^d}\right] \geq c_0 > 0 \tag{2.16}$$

holds for $\theta = \theta_1$ and some $\mu, c_0 > 0$. Then we have for some $C = C(B, \mu, c_0, \mathcal{X}, g, h, f_{min}) > 0$,

$$\|\theta_1 - \theta_2\|_{L^2} \leq C\|u_{f_{\theta_1}} - u_{f_{\theta_2}}\|_{H^2}. \tag{2.17}$$

Proof. We will use throughout that $u_{f_\theta} \in H^{\beta+1}(\mathcal{X})$ defines an element of $C^2(\mathcal{X})$ in view of (6.4) and (6.11) with $V = 0$. Define the operator

$$h \mapsto T_\theta(h) := \nabla \cdot (h\nabla u_{f_\theta}), \ \ T_\theta : H^1_c(\mathcal{X}) \to L^2(\mathcal{X}).$$

Lemma 2.1.6. *Under the hypotheses of Proposition 2.1.5 with $\theta = \theta_1$, we have*

$$\|T_\theta(h)\|_{L^2} = \|\nabla \cdot (h \nabla u_{f_\theta})\|_{L^2} \geq c\|h\|_{L^2}$$

for all $h \in H^1_c(\mathcal{X})$ and some constant $c = c(\mu, U, c_0) > 0$, where U is from (2.9).

Proof. Let us write $u_\theta = u_{f_\theta}$ in the proof. Applying the divergence theorem (6.7) to any $v \in C^1(\mathcal{X})$ vanishing at $\partial\mathcal{X}$ gives

$$\langle \Delta u_\theta, v^2 \rangle_{L^2} + \frac{1}{2}\langle \nabla u_\theta, \nabla(v^2)\rangle_{L^2} = \frac{1}{2}\langle \Delta u_\theta, v^2\rangle_{L^2}.$$

Consider first $h \in C^\infty(\mathcal{X})$ of compact support in $\mathcal{X}$ (and $h \neq 0$ without loss of generality). Scaling h by integrating factors we set $v = e^{-\mu u_\theta} h$ with $\mu > 0$ to be chosen and obtain

$$\frac{1}{2}\int_{\mathcal{X}} \nabla(v^2) \cdot \nabla u_\theta = -\int_{\mathcal{X}} \mu |\nabla u_\theta|^2_{\mathbb{R}^d} v^2 + \int_{\mathcal{X}} v e^{-\mu u_\theta} \nabla h \cdot \nabla u_\theta,$$

so that by the Cauchy-Schwarz inequality

$$\begin{aligned}
\left|\int_{\mathcal{X}} \Big(\frac{1}{2}\Delta u_\theta + \mu|\nabla u_\theta|^2_{\mathbb{R}^d}\Big) v^2 \right| &= \left|\langle(\Delta u_\theta + \mu|\nabla u_\theta|^2_{\mathbb{R}^d}), v^2\rangle_{L^2} + \frac{1}{2}\langle \nabla u_\theta, \nabla(v^2)\rangle_{L^2}\right| \\
&= \left|\langle h\Delta u_\theta + \nabla h \cdot \nabla u_\theta, h e^{-2\mu u_\theta}\rangle_{L^2}\right| \\
&\leq \bar{\mu}\|\nabla \cdot (h\nabla u_\theta)\|_{L^2}\|h\|_{L^2} \qquad (2.18)
\end{aligned}$$

for $\bar{\mu} = \exp(2\mu\|u_\theta\|_\infty)$. We next lower bound the multipliers of v^2 in l.h.s. of (2.18): By (2.16)

$$\left|\int_{\mathcal{X}} \Big(\frac{1}{2}\Delta u_\theta + \mu|\nabla u_\theta|^2_{\mathbb{R}^d}\Big) v^2\right| \geq c_0 \int_{\mathcal{X}} v^2$$

and combining this with (2.18) we deduce for appropriate $c' > 0$

$$\|\nabla \cdot (h\nabla u_\theta)\|_{L^2}\|h\|_{L^2} \geq c'\|v\|^2_{L^2} \gtrsim \|h\|^2_{L^2}$$

which proves the lemma for all $h \in C^\infty(\mathcal{X})$ of compact support. If a sequence h_n of such functions converges in $H^1(\mathcal{X})$-norm to some h as $n \to \infty$, then h_n and using (6.3) also $T_\theta(h_n)$ converge in L^2 to h and $T_\theta(h)$, respectively, and hence the inequality of the lemma is preserved after taking limits, so that the result extends to the completion $H^1_c(\mathcal{X})$ of $C^\infty_c(\mathcal{X})$ for the $\|\cdot\|_{H^1}$-norm. □

To deduce Proposition 2.1.5, we first notice that by the mean value theorem, for $\tilde{\theta}(x) \in [\theta_1(x), \theta_2(x)]$,

$$f_{\theta_1} - f_{\theta_2} = e^{\theta_1} - e^{\theta_2} = e^{\tilde{\theta}}(\theta_1 - \theta_2)$$

and since the θ_i are bounded in C^1 they and then also $\tilde{\theta}$ are uniformly bounded by a constant depending only on B, whence

$$\|\theta_1 - \theta_2\|_{L^2} \leq c(B)\|f_{\theta_1} - f_{\theta_2}\|_{L^2}. \tag{2.19}$$

Then let us write $h = f_{\theta_1} - f_{\theta_2}$ which defines an element of $H_0^1(\mathcal{X}) = H_c^1(\mathcal{X})$ under the hypotheses maintained. By (1.2) we have

$$\nabla \cdot (h \nabla u_{\theta_1}) = \nabla \cdot (f_{\theta_2} \nabla (u_{\theta_2} - u_{\theta_1})) = \mathscr{L}_{f_{\theta_2}}(u_{\theta_2} - u_{\theta_1})$$

and hence for constants on the r.h.s. depending only on an upper bound for $\|f_{\theta_2}\|_{C^1} \lesssim c(B) < \infty$ we obtain

$$\|\nabla \cdot (h \nabla u_{\theta_1})\|_{L^2} \lesssim \|u_{\theta_2} - u_{\theta_1}\|_{H^2}.$$

By Lemma 2.1.6 the l.h.s. is lower bounded by a constant multiple of $\|h\|_{L^2} = \|f_{\theta_1} - f_{\theta_2}\|_{L^2}$, so that the result follows. □

Condition (2.16) can be verified for a large class of constellations of h, g in (1.2) with $\mathscr{L}_f$ from (1.3). It clearly holds whenever the gradient of u_θ does not vanish (for examples see Ex. 2.4.2). More generically if $h = 0$ and the 'initial temperatures' in (1.5) are strictly positive, $g \geq g_{min} > 0$, then

$$0 < g_{min} \leq g = f_\theta \Delta u_{f_\theta} + \nabla f_\theta \cdot \nabla u_{f_\theta} \quad \text{on } \mathcal{X}, \tag{2.20}$$

so that either $\Delta u_{f_\theta} \geq g_{min}/(2\|f_\theta\|_\infty)$ or $|\nabla u_{f_\theta}(x)|_{\mathbb{R}^d} \geq g_{min}/(2\|f_\theta\|_{C^1})$ has to hold on $\mathcal{X}$, which allows one to check Condition (2.16).

To conclude, the verification of Condition 2.1.4 for $\mathcal{R} = H^\beta(\mathcal{X}), \beta > 1 + d/2$, now follows from Proposition 2.1.5, an elliptic regularity estimate for solutions u_{f_θ} of (1.2), and interpolation. This is summarised in the final result of this subsection.

Proposition 2.1.7. *Let $\theta, \theta_0 \in H^\beta(\mathcal{X})$ for some $\beta > 1 + d/2$ satisfy $\theta = \theta_0 = 0$ on $\partial\mathcal{X}$ and let $B \geq \max(\|\theta\|_{H^\beta}, \|\theta_0\|_{H^\beta})$. Let $\mathscr{G}(\theta) = u_\theta$ be the forward map from (2.8) and suppose further (2.16) holds for $\theta = \theta_0$ and some μ, c_0. Then there exists a constant $c = c(B, \mu, c_0, \mathcal{X}, g, h, f_{min}, \beta)$ such that*

$$\|\theta - \theta_0\|_{L^2} \leq c\|\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\|_{L^2}^\eta, \quad \eta = \frac{\beta - 1}{\beta + 1}, \tag{2.21}$$

in particular Condition 2.1.4 holds with this choice of η and $L' = L'(B)$ if the regularisation norm $\|\cdot\|_{\mathcal{R}}$ dominates the H^β-norm.

Proof. Since the H^β norms bound the C^1-norms by (6.4) we deduce from Proposition 2.1.5 and (6.5) that

$$\|\theta - \theta_0\|_{L^2} \lesssim \|u_\theta - u_{\theta_0}\|_{H^2} \lesssim \|u_\theta - u_{\theta_0}\|_{L^2}^{(\beta-1)/(\beta+1)} \|u_\theta - u_{\theta_0}\|_{H^{\beta+1}}^{2/(\beta+1)}$$

with constants only depending on permitted quantities. The last factor is uniformly bounded in view of the elliptic regularity estimate Proposition 6.1.5 with $\gamma = f_\theta, f_{\theta_0}, V = 0$, so that the result follows. □

2.2 Regularisation with Gaussian process priors

In this section $\mathcal{Z}$ will be a bounded domain in $\mathbb{R}^d$ with smooth boundary. [For $\mathcal{Z}$ equal to a d-dimensional Torus, the proofs that follow work as well if one considers periodic Sobolev spaces.] We will apply Theorem 1.3.2 to prior measures Π that arise from Gaussian processes $(X(z) : z \in \mathcal{Z})$ indexed by the domain $\mathcal{Z}$. That is, Π will be an infinite-dimensional Gaussian distribution that is the law of a random field taking values in a closed *linear* subspace Θ of the Hilbert space $L^2_\zeta(\mathcal{Z})$, see Section 6.2.1 for a review of these concepts.

Normed linear subspaces $\mathcal{R} \subseteq \Theta$ where the measure Π is supported – i.e.,

$$\Pi(\mathcal{R}) = \Pi\big(\theta \in \Theta : \|\theta\|_{\mathcal{R}} < \infty\big) = 1,$$

will be shown to serve as choices for the 'regularisation spaces' in Conditions 2.1.1 and 2.1.4 – they often encode 'path regularity' of the Gaussian process on its index set $\mathcal{Z}$. Further features of these prior measures are related to so-called *reproducing kernel Hilbert spaces* (RKHS) $\mathcal{H}$ of the prior. These are also subspaces of Θ but do not (in infinite dimensions) support Π. Rather they describe geometric 'covariance-related' properties of Π. In the theory that follows we can think of $\mathcal{R} = H^\beta(\mathcal{Z})$ and $\mathcal{H} = H^\alpha(\mathcal{Z})$ with $\alpha > \beta + d/2$, perhaps endowed with further boundary conditions (via $H^\beta_c(\mathcal{Z})$ or $H^\beta_0(\mathcal{Z})$ spaces, see Subsection 6.1.1).

Condition 2.2.1. *Let Π' be a centred Gaussian Borel probability measure on the linear space $\Theta \subseteq L^2_\zeta(\mathcal{Z}, W)$ with RKHS $\mathcal{H}$. Suppose further that $\Pi'(\mathcal{R}) = 1$ for some separable normed linear subspace $(\mathcal{R}, \|\cdot\|_{\mathcal{R}})$ of Θ.*

In Section 6.2.1 we review some standard examples of Gaussian processes satisfying Condition 2.2.1 (see also Ex. 2.4.3). The prior $\Pi = \Pi_N$ arises from the base prior Π' as the law $\mathcal{L}(\theta)$ after 'rescaling'

$$\theta = N^{-d/(4\alpha+4\kappa+2d)}\theta', \quad \theta' \sim \Pi', \tag{2.22}$$

for 'regularity' parameter $\alpha > 0$ to be chosen and where κ is the 'forward smoothing degree' of $\mathscr{G}$ from (2.4). In other words we shrink the random Gaussian series θ' towards zero – quantitatively this entails (in our context, *essential*) extra regularisation of the posterior measure since one can heuristically (say if $\mathcal{H}$ is finite-dimensional) think of the 'reweighting' factor in (1.14) as being of Gaussian density form

$$d\Pi(\cdot) \propto \exp\Big\{-\frac{N^{d/(2\alpha+2\kappa+d)}\|\cdot\|^2_{\mathcal{H}}}{2}\Big\},$$

thus penalising large values of $\|\theta\|_{\mathcal{H}}$ more than one would do without re-scaling. While the catonic (subjective) Bayesian may wish to avoid such N-dependent priors, penalties of this type are widely used in the analysis of regularised least squares procedures (e.g., p.174 in [125]) and also in the Bayesian context [128].

Theorem 2.2.2. *Suppose Condition 2.1.1 holds for the forward map $\mathscr{G}$, some $\kappa \geq 0, U < \infty$, regularisation space $\mathcal{R}$ and bounded domain $\mathcal{Z} \subset \mathbb{R}^d$ with smooth boundary. Let the prior Π' satisfy Condition 2.2.1 for this choice of $\mathcal{R}$ and suppose its RKHS $\mathcal{H}$ satisfies the continuous imbedding*

$$\mathcal{H} \subset H_c^\alpha(\mathcal{Z}) \text{ if } \kappa \geq 1/2 \quad \text{or } \mathcal{H} \subset H^\alpha(\mathcal{Z}) \text{ if } \kappa < 1/2,$$

and denote by Π_N the law of the rescaled prior from (2.22). If $\theta_0 \in \mathcal{H} \cap \mathcal{R}$, then the hypotheses (1.32), (1.33) and (1.34) in Theorem 1.3.2 hold for some $A > 0$, for sequence

$$\delta_N = N^{-(\alpha+\kappa)/(2\alpha+2\kappa+d)}, \tag{2.23}$$

and any B if the 'regularisation sets' are chosen as

$$\Theta_N = \Big\{\theta \in \mathcal{R} : \theta = \theta_1 + \theta_2, \|\theta_1\|_{(H^\kappa)^*} \leq M\delta_N, \|\theta_2\|_{\mathcal{H}} \leq M, \|\theta\|_{\mathcal{R}} \leq M\Big\}, \tag{2.24}$$

for $M = M(B)$ sufficiently large. In particular (1.35) holds with $V = U < \infty$.

Proof. The proof is organised in several steps: The small ball computation in ii) determines the sequence δ_N, and i) and iii) combined verify the deviation bound (1.33), while iv) checks the complexity condition (1.34). We note that the proof in fact works for any subset $\bar{\Theta}_N$ of Θ_N as long as $\Pi(\bar{\Theta}_N^c) \leq e^{-BN\delta_N^2}$ holds for some $B > A + 2$. We use throughout that $N^{d/(4\alpha+4\kappa+2d)} = \sqrt{N}\delta_N$ for our choice of δ_N.

i) As $\mathcal{R}$ is separable the Hahn-Banach theorem implies that its norm can be represented as

$$\|\theta'\|_{\mathcal{R}} = \sup_{T \in \mathcal{T}} |T(\theta')|, \quad \theta' \in \mathcal{R},$$

where $\mathcal{T}$ is a countable family of continuous linear forms on $(\mathcal{R}, \|\cdot\|_{\mathcal{R}})$. In particular for $\theta' \sim \Pi'$ the collection of random variables $\{T(\theta') : T \in \mathcal{T}\}$ defines a centred Gaussian process with countable index set and $\Pr(\|\theta'\|_{\mathcal{R}} = \sup_{T\in\mathcal{T}} |T(\theta')| < \infty) = 1$ by hypothesis. Thus Fernique's theorem (Theorem 2.1.20 in [61]) implies initially that $E\|\theta'\|_{\mathcal{R}} \leq D$ for some constant D depending only on the base prior Π', and then gives the bound

$$\begin{aligned}\Pi(\|\theta\|_{\mathcal{R}} > M) &= \Pi'(\|\theta'\|_{\mathcal{R}} > M\sqrt{N}\delta_N) \\ &\leq \Pi'(\|\theta'\|_{\mathcal{R}} - E\|\theta'\|_{\mathcal{R}} > M\sqrt{N}\delta_N/2) \leq e^{-cM^2N\delta_N^2} \leq \frac{1}{2}e^{-BN\delta_N^2}\end{aligned} \tag{2.25}$$

for all M large enough (and since $\sqrt{N}\delta_N \geq 1$).

ii) We next address the small ball computation for (1.32) with sets $\mathcal{B}_N$ from (1.31). Since θ_0 is assumed to belong to $\mathcal{R}$ we have for all M and $\bar{M} = \bar{M}(M)$ large enough that

$$\|\theta - \theta_0\|_{\mathcal{R}} \leq M \Rightarrow \|\theta\|_{\mathcal{R}} \leq M + \|\theta_0\|_{\mathcal{R}} \equiv \bar{M}$$

and by (2.3) in Condition 2.1.1 then also $\|\mathscr{G}(\theta)\|_\infty \leq U$ for some constant $U = U_{\mathscr{G}}(\bar{M})$. Next, using also (2.4) in Condition 2.1.1, Corollary 2.6.18 in [61] and the Gaussian correlation inequality Theorem 6.2.2 below we have

$$\begin{aligned}
&\Pi(\theta : \|\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\|_{L^2} \leq \delta_N, \|\mathscr{G}(\theta)\|_\infty \leq U) \\
&\geq \Pi(\theta : \|\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\|_{L^2} \leq \delta_N, \|\theta - \theta_0\|_{\mathcal{R}} \leq M) \\
&\geq \Pi(\theta : \|\theta - \theta_0\|_{(H^\kappa)^*} \leq \delta_N / L_{\mathscr{G}}(\bar{M}), \|\theta - \theta_0\|_{\mathcal{R}} \leq M) \\
&\geq e^{-\frac{1}{2}N\delta_N^2\|\theta_0\|_{\mathcal{H}}^2} \Pi(\theta : \|\theta\|_{(H^\kappa)^*} \leq \delta_N / L_{\mathscr{G}}(\bar{M})) \cdot \Pi(\|\theta\|_{\mathcal{R}} \leq M). \qquad (2.26)
\end{aligned}$$

The last probability equals

$$\Pi'(\|\theta'\|_{\mathcal{R}} \leq M\sqrt{N}\delta_N) = 1 - \Pi'(\|\theta'\|_{\mathcal{R}} > M\sqrt{N}\delta_N) \geq 1 - e^{-cMN\delta_N^2} \geq 1/2$$

by virtue of (2.25) in step i) (and for M large enough). So it remains to lower bound the middle factor in (2.26) which equals

$$\Pi'(\theta' : \|\theta'\|_{(H^\kappa)^*} \leq \sqrt{N}\delta_N^2 / L_{\mathscr{G}}(\bar{M})).$$

We use (6.25) to the effect that a ball $h_c^\alpha(r)$ of radius r in $\mathcal{H} \subset H_c^\alpha$ has log-ϵ-covering numbers bounded as

$$\log N\big(h_c^\alpha(r), \|\cdot\|_{(H^\kappa)^*}, \epsilon\big) \lesssim \epsilon^{-d/(\alpha+\kappa)}, \ 0 < \epsilon < r, \qquad (2.27)$$

and this remains valid for $\kappa < 1/2$ also with H_c^α replaced by H^α as discussed after (6.25). Hence an application of Theorem 6.2.1 in the ambient Banach space $B = (H^\kappa(\mathcal{Z}))^* \supset L^2(\mathcal{Z})$ supporting Π' gives, after some basic calculations,

$$\Pi'(\theta' : \|\theta'\|_{(H^\kappa)^*} \leq \sqrt{N}\delta_N^2 / L_{\mathscr{G}}(\bar{M})) \geq e^{-\bar{a}N\delta_N^2} \qquad (2.28)$$

for some $\bar{a} > 0$. Overall the r.h.s. in (2.26) is lower bounded by

$$\frac{1}{2}\exp\big\{-(\bar{a} + \|\theta_0\|_{\mathcal{H}}^2/2)N\delta_N^2\big\} \geq e^{-AN\delta_N^2}$$

for $A = \bar{a} + \|\theta_0\|_{\mathcal{H}}^2/2 > 0$.

iii) We next show (1.33). By Step i) it suffices to prove

$$\Pi(\theta : \theta = \theta_1 + \theta_2 : \|\theta_1\|_{(H^\kappa)^*} \leq M\delta_N, \|\theta_2\|_{\mathcal{H}} \leq M) \geq 1 - \frac{1}{2}\exp\{-BN\delta_N^2\}$$

for M large enough. We can ignore the 1/2 factor by increasing the constant B. For the rescaled prior measures this requires a lower bound for

$$\Pi'\big(\theta' : \theta' = \theta_1' + \theta_2', \ \|\theta_1'\|_{(H^\kappa)^*} \leq M\sqrt{N}\delta_N^2, \|\theta_2'\|_{\mathcal{H}} \leq M\sqrt{N}\delta_N\big).$$

Using Theorem 6.2.1 as before (2.28) we deduce that for some $c > 0$

$$-\log \Pi'(\theta' : \|\theta'\|_{(H^\kappa)^*} \le M\sqrt{N}\delta_N^2) \le c(M\sqrt{N}\delta_N^2)^{-\frac{2d}{2\alpha+2\kappa-d}}$$

so that taking any $M > (B/c)^{-(2\alpha+2\kappa-d)/(2d)}$ implies

$$-\log \Pi'(\theta' : \|\theta'\|_{(H^\kappa)^*} \le M\sqrt{N}\delta_N^2) \le B(\sqrt{N}\delta_N^2)^{-\frac{2d}{2\alpha+2\kappa-d}} = BN\delta_N^2. \qquad (2.29)$$

Next, denote

$$B_N = -2\Phi^{-1}(e^{-BN\delta_N^2})$$

where Φ is the standard normal cumulative distribution function. Then by standard inequalities for Φ^{-1} (e.g., Lemma K.6 in [59]) we have $B_N \simeq \sqrt{BN}\delta_N$ as $N \to \infty$, so that for $M \ge \sqrt{B}$

$$\begin{aligned}\Pi'(\theta' : \theta' = \theta_1' + \theta_2',\ \|\theta_1'\|_{(H^\kappa)^*} \le M\sqrt{N}\delta_N^2, \|\theta_2'\|_{\mathcal{H}} \le M\sqrt{N}\delta_N)\\ \ge \Pi'(\theta' : \theta' = \theta_1' + \theta_2',\ \|\theta_1'\|_{(H^\kappa)^*} \le M\sqrt{N}\delta_N^2, \|\theta_2'\|_{\mathcal{H}} \le B_N).\end{aligned}$$

By (2.29) and the isoperimetric inequality for Gaussian measures [61, Theorem 2.6.12], the last probability is then lower bounded by

$$\Phi(\Phi^{-1}[\Pi_N'(\|\theta'\|_{(H^\kappa)^*} \le M\sqrt{N}\delta_N^2)] + B_N) \ge \Phi(\Phi^{-1}[e^{-BN\delta_N^2}] + B_N) = 1 - e^{-BN\delta_N^2},$$

concluding the proof.

iv) It remains to prove (1.34). By definition of the set Θ_N and (2.4) from Condition 2.1.1, it suffices to construct a $\bar{m}\delta_N/4$-covering in $(H^\kappa)^*$-distance of a ball in $\mathcal{H}$ intersected with a ball in $\mathcal{R}$, for all $\bar{m} = \bar{m}(M, L)$ large enough. The bound required in (1.34) then follows directly from (2.27) with $\epsilon = (\bar{m}/4)\delta_N$ and $\bar{m}$ large enough. [The case $\kappa < 1/2$ again follows similarly, cf. after (6.25)]. □

The last theorem applies readily to choices for Π' discussed in Section 6.2.1. For high-dimensional 'sieved' priors expressed in appropriate basis functions of the Dirichlet-Laplacian, the proofs require minor adjustments and also permit some refinements regarding the 'regularisation sets' Θ_N – see Ex. 2.4.3. We also note that when $\kappa = 0$, it suffices to require Condition 2.1.1 with the stronger $\|\cdot\|_\infty$-norm replacing the $L^2 = H^0$ norm on the r.h.s. in (2.4), see Exercise 2.4.4.

2.3 Convergence of posterior measure and mean

From the preceding developments we can now state the following posterior contraction rate result in the inverse problem at hand.

Theorem 2.3.1. *Suppose the Gaussian process prior Π_N and forward map $\mathscr{G}$ satisfy the conditions of Theorem 2.2.2 with Θ_N as in (2.24), some regularisation norm $\mathcal{R}$, $\delta_N = N^{-(\alpha+\kappa)/(2\alpha+2\kappa+d)}$ and some $\kappa \geq 0$. If $\Pi(\cdot|(Y_i, X_i)_{i=1}^N) = \Pi(\cdot|D_N)$ is the posterior distribution from (1.14) arising from observations in the model (1.11), then for all $b > 0$ we can choose m large enough such that as $N \to \infty$,*

$$P_{\theta_0}^N\Big(\Pi_N\big(\theta \in \Theta_N, \|\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\|_{L^2_\lambda(\mathcal{X},V)} \leq m\delta_N|D_N\big) \leq 1 - e^{-bN\delta_N^2}\Big) \to 0. \tag{2.30}$$

Moreover, assuming also Condition 2.1.4 for the present choice of $\mathcal{R}$ and some $\eta, L' > 0$, we deduce that

$$P_{\theta_0}^N\Big(\Pi_N\big(\theta \in \Theta_N, \|\theta - \theta_0\|_{L^2_\zeta} \leq L'(m\delta_N)^\eta|D_N\big) \leq 1 - e^{-bN\delta_N^2}\Big) = o(1) \tag{2.31}$$

Proof. From Theorems 1.3.2 and 2.2.2 we deduce directly the contraction rate (2.33) on the 'forward level'. Next Condition 2.1.4 implies the set inclusion

$$\{\theta \in \Theta_N, \|\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\|_{L^2_\lambda(\mathcal{X},V)} \leq m\delta_N\} \subset \{\|\theta - \theta_0\|_{L^2_\zeta} \leq L'(m\delta_N)^\eta\}$$

so that (2.31) follows directly from (2.33). □

To construct a concrete statistical estimator (or 'algorithm'), we extract a specific feature of the posterior distribution such as the mean $E^\Pi[\theta|D_N]$ or posterior mode (maximiser of the posterior surface). The former is well defined as a Bochner integral and a natural Bayesian estimator, not the least since we can form ergodic averages of the MCMC outputs to numerically approximate it. The following theorem shows how the contraction rate from (2.3.1) translates into a convergence rate for $E^\Pi[\theta|D_N]$.

Theorem 2.3.2. *Under the conditions of (2.31) in Theorem 2.3.1, we have that*

$$\|E^\Pi[\theta|D_N] - \theta_0\|_{L^2_\zeta} = O_{P_{\theta_0}^N}(\delta_N^\eta). \tag{2.32}$$

Proof. Let us write $L^2 = L^2_\zeta$ in this proof and set $\eta_N = L'(m\delta_N)^\eta$, with constant m to be chosen large enough. Then by the inequalities of Jensen and Cauchy-Schwarz

$$\begin{aligned}\|E^\Pi[\theta|D_N] - \theta_0\|_{L^2} &\leq E^\Pi[\|\theta - \theta_0\|_{L^2}|D_N] \\ &\leq \eta_N + E^\Pi[\|\theta - \theta_0\|_{L^2} 1\{\|\theta - \theta_0\|_{L^2} > \eta_N\}|D_N] \\ &\leq \eta_N + [E^\Pi[\|\theta - \theta_0\|_{L^2}^2|D_N]^{1/2}\Pi(\|\theta - \theta_0\|_{L^2} > \eta_N|D_N)^{1/2}\end{aligned}$$

and we now show that last term is $O_{P_{\theta_0}^N}(\eta_N)$ to prove the theorem. We recall the sets A_N from (1.37) which satisfy $P_{\theta_0}^N(A_N) \to 1$ as $N \to \infty$. Then using (2.31), Markov's inequality, Fubini's theorem, (1.39) and that the Gaussian measure Π' is supported in L^2 and hence integrates $\|\cdot\|_{L^2}^2$ to a finite constant,

$$\begin{aligned}
&P_{\theta_0}^N\Big(E^\Pi[\|\theta-\theta_0\|_{L^2}^2|D_N]\times\Pi(\|\theta-\theta_0\|_{L^2}>\eta_N|D_N)>\eta_N^2\Big)\\
&\le P_{\theta_0}^N\Big(E^\Pi[\|\theta-\theta_0\|_{L^2}^2|D_N]e^{-bN\delta_N^2}>\eta_N^2\Big)+o(1)\\
&\le P_{\theta_0}^N\Big(e^{-bN\delta_N^2}\frac{\int\|\theta-\theta_0\|_{L^2}^2e^{\ell_N(\theta)-\ell_N(\theta_0)}d\Pi(\theta)}{\int e^{\ell_N(\theta)-\ell_N(\theta_0)}d\Pi(\theta)}>\eta_N^2, A_N\Big)+o(1)\\
&\le e^{(A+2-b)N\delta_N^2}\eta_N^{-2}\int\|\theta-\theta_0\|_{L^2}^2E_{\theta_0}^N[e^{\ell_N(\theta)-\ell_N(\theta_0)}]d\Pi(\theta)+o(1)\\
&\le e^{(A+2-b)N\delta_N^2}\eta_N^{-2}\int\|\theta-\theta_0\|_{L^2}^2d\Pi(\theta)+o(1)\to_{N\to\infty}0,
\end{aligned}$$

for m and then b large enough. □

Putting all what precedes together we can obtain posterior contraction theorems for the non-linear inverse problem arising from the model PDE (1.2) with $\mathscr{L}_f$ from (1.3) or (1.4). We formulate a detailed result for Darcy's problem and leave the Schrödinger case to Ex. 2.4.1.

Recall from (2.20) that simple sufficient conditions for the hypothesis (2.16) in the following theorem can be given in terms of the source function g in (1.2), and that natural priors satisfying the required hypotheses exist (Section 6.2.1).

Theorem 2.3.3 (Bayes solution of Darcy's problem)**.** *Consider a Gaussian process prior Π' verifying Condition 2.2.1 with $\Theta=\mathcal{R}=H_c^\beta(\mathcal{X})$ for some $\beta>1+d/2$ and RKHS $\mathcal{H}\subset H_c^\alpha(\mathcal{X})$ for some $\alpha>\beta+d/2$. Let the rescaled prior Π_N arise from (2.22) for some $0\le\kappa\le 1$. Consider the forward map $\mathscr{G}:\Theta\to L_\lambda^2(\mathcal{X})$ from (2.8) and assume that (2.16) is satisfied for $\theta=\theta_0\in\mathcal{H}$. Let $\Pi(\cdot|(Y_i,X_i)_{i=1}^N)$ be the posterior distribution from (1.14) arising from observations in the model (1.11). Let $\delta_N=N^{-(\alpha+\kappa)/(2\alpha+2\kappa+d)}$. Then for all $b>0$ we can choose m large enough such that as $N\to\infty$,*

$$P_{\theta_0}^N\Big(\Pi_N\big(\theta,\|\mathscr{G}(\theta)-\mathscr{G}(\theta_0)\|_{L_\lambda^2}\le m\delta_N,\|\theta\|_{H^\beta}\le m|D_N\big)\le 1-e^{-bN\delta_N^2}\Big)=o(1) \tag{2.33}$$

Moreover, for $\eta=(\beta-1)/(\beta+1)$ and constant $M>0$ large enough, we also have

$$P_{\theta_0}^N\Big(\Pi_N\big(\theta:\|\theta-\theta_0\|_{L_\lambda^2}>M\delta_N^\eta|D_N\big)\ge e^{-bN\delta_N^2}\Big)=o(1)$$

as $N\to\infty$, and likewise the posterior mean satisfies

$$\|E^\Pi[\theta|D_N]-\theta_0\|_{L_\lambda^2}=O_{P_{\theta_0}^N}(\delta_N^\eta). \tag{2.34}$$

Proof. Proposition 2.1.3 verifies Conditions 2.1.1 for $\kappa = 1$ and then for any $0 \leq \kappa \leq 1$. Moreover for all $\theta \in \Theta = H^\beta_c \subset H^\beta_0$ (see Section 6.1.1) the boundary values $(f_\theta)_{|\partial\mathcal{X}} = 1 + f_{min}$ co-incide at $\partial\mathcal{X}$, so that Proposition 2.1.7 verifies Condition 2.1.4. The result then follows from Theorems 2.3.1 and 2.3.2. □

One shows easily that the same convergence rate toward f_{θ_0} is inherited by the induced estimates $f_\theta = f_{min} + e^\theta, \theta \sim \Pi(\cdot|(Y_i, X_i)_{i=1}^N)$ of the conductivity, since e^θ is Lipschitz on bounded sets of H^β. Inspection of the proof further shows that when $0 \leq \kappa < 1/2$, one can replace H^β_c, H^α_c by H^β_0, H^α_0, respectively, in the hypotheses of the preceding theorem. The preceding theorem shows that the posterior contracts towards θ_0 at a rate as close to $1/\sqrt{N}$ as desired in the 'smooth case' where $\alpha, \beta \to \infty$, which is near-optimal in this situation. A discussion of more general optimality considerations can be found in the notes below.

Remark 2.3.4. [Relationship to other inverse problems.] A classical non-linear inverse problem (the Calderón problem) arising with the diffusion operator $\mathscr{L}_f$ from (1.3) was discussed already briefly in Section 1.1.1, where only *boundary measurements* of u_f are available. We refer to [1] for an account of the precise forms of statistical measurement models one can consider. As mentioned earlier the resulting forward map $\mathscr{G}$ is injective [93, 120] even in absence of interior measurements, but it is also 'severely ill-posed' in the sense that the recovery guarantees are at best of logarithmic order $(\log N)^{-a}, a > 0$, in natural statistical measurement settings and with parameter spaces for f paralleling those in Theorem 2.3.3. A proof of this fact as well as many more relevant references can be found in [1]. We can thus conclude from the previous theorem that the availability of 'interior' measurements of the solutions u_f of (1.2) in Darcy's problem substantially accelerates these convergence rates to algebraic ones (i.e., N^{-a} for some $a > 0$). Note however that in other settings where only boundary measurements ('scattering data') are available, algebraic rates can be achieved – see [91] for the non-linear inverse problem with non-Abelian X-ray transforms from (1.1), with proofs that follow the same lines as those laid out in this chapter. The information-theoretic complexity of statistical recovery rates thus depends on possibly subtle interactions between the types of measurements one takes and the underlying partial differential equation. For data assimilation problems discussed in Subsection 1.1.3, see [95, 100].

2.4 Notes

2.4.1 Exercises

Exercise 2.4.1. Consider the Schrödinger forward map $\mathscr{G}$ from (2.5) and let $\theta_0 \in H^\xi$ for some $\xi > 2 + d/2$. Show that $\mathscr{G}$ satisfies the forward regularity

Condition 2.1.1 with $\kappa = 2$ and $\mathcal{R} = H^\xi$. [Hint: adapt the arguments from (2.11) and below it, or see the proof of Theorem 12 in [101].] Show further that $\mathscr{G}$ satisfies the stability Condition 2.1.4 with this choice of $\mathcal{R}$ and $\eta = \xi/(2+\xi)$. [Hint: Track the constants in (2.13) and then use interpolation and elliptic regularity as in the proof of Proposition 2.1.7; alternatively see Lemma 28 in [101].] Then prove an analogue of Theorem 2.3.3 with contraction rates $\delta_N^\eta, \delta_N = N^{-(\alpha+2)/(2\alpha+4+d)}$.

Exercise 2.4.2. For $\mathcal{L}_f$ from (1.3) with $f = 1$ (the standard Laplacian), consider solutions u_f to (1.2) with $g = 2$ identically and $h = (|\cdot|^2_{\mathbb{R}^d} - 1)/d$. Show that if $\mathcal{X}$ is a domain separated away from the origin, then $|\nabla u_1| \geq c > 0$ on $\mathcal{X}$. Show by perturbation that this lower bound extends to ∇u_f whenever $\|f - 1\|_{H^\beta} < \epsilon$, ϵ small enough, and $\beta > 1 + d$. Then verify Condition (2.16) for such $f = f_\theta$.

Exercise 2.4.3. For eigen-pairs $(\lambda_j, e_j) \in (0, \infty) \times H^1_0$ of the Dirichlet-Laplacian $-L_{1,0}$ from (6.9) and δ_N as in (2.23) with $\kappa = 0$, consider a prior $\Pi_N = \mathrm{Law}(\theta)$ as

$$\theta = \frac{1}{\sqrt{N}\delta_N} \sum_{j=1}^{D} g_j \lambda_j^{-\alpha/2} e_j, \quad D \lesssim N\delta_N^2, \;\; \alpha > 1 + d/2.$$

For $\theta \in \Theta$ denote its projection onto the linear span E_D of $\{e_j : j \leq D\}$ by θ_D. Suppose for $\mathscr{G}$ satisfying Condition 2.1.1 and ground truth $\theta_0 \in H^\alpha_0$ we have

$$\|\mathscr{G}(\theta_0) - \mathscr{G}(\theta_{0,D})\|_{L^2} \leq \delta_N/2. \tag{2.35}$$

Show that the conclusions of Theorem 2.2.2 hold true with regularisation sets

$$\Theta_N \subset \{\theta \in H^\alpha_0 \cap E_D : \|\theta\|_{H^\alpha} \leq M\} \tag{2.36}$$

and that hence Theorem 2.3.3 holds with β replaced by α for this prior. [Hint: If Π' is the law of the infinite series from (6.45) then its RKHS $\mathcal{H}$ equals $\tilde{H}^\alpha \subset H^\alpha_0$. We also have $\Pi'(\mathcal{R}) = 1$ for $\mathcal{R} = H^\beta_0$ just as in the last step of the proof of Theorem 6.2.3. We can then follow the steps of the proof of Theorem 2.2.2 for this choice of $\mathcal{H}, \mathcal{R}, \kappa = 0$, and with Π_N the law of $\theta = \theta'_D/(\sqrt{N}\delta_N)$, $\theta' \sim \Pi'$. One uses that the projection onto E_D does not increase $\mathcal{H}, \mathcal{R}$-norms, and in the small ball calculation (2.26) we use (2.35). On the sets Θ_N from (2.24) we can then use the definition of the $\tilde{H}^\alpha$-norm in (6.15), Weyl's law (6.14) and Parseval's identity, to estimate

$$\|\theta_1\|_{H^\alpha} \lesssim \|\theta_1\|_{\tilde{H}^\alpha} \lesssim D^{\alpha/d} \|\theta_1\|_{L^2} \leq (N\delta_N^2)^{\alpha/d} \delta_N \leq M.]$$

Exercise 2.4.4. Prove Theorem 2.2.2 for the case $\kappa = 0$ and under the weaker hypothesis where $\|\theta_1 - \theta_2\|_\infty$ replaces $\|\theta_1 - \theta_2\|_{(H^0)^*}$ on the r.h.s. of (2.4) in Condition 2.1.1. [Hint: The proof is the same, but instead of (2.27), one uses the entropy bound (4.184) in [61] and applies Theorem 6.2.1 in the Banach space $L^\infty(\mathcal{Z})$.)

2.4.2 Remarks and comments

Conditions such as 2.1.1 and 2.1.4 have been used implicitly or explicitly in the statistical analysis of non-linear inverse problems for a while, we mention [94, 98, 101, 134] among others. Their verification for the elliptic PDE models here is fairly routine but does require some attention to the exact dependence of the constants on the parameter space employed: Proposition 2.1.5 is from [101] – see also [70, 96] – but results of this flavour go back as far as [111], with a recent reference on the topic being [24]. The main ideas combine with more complex non-linear inverse problems where such stability estimates are often also available (but more difficult to obtain), see, e.g., [91, 104], [21, 23] and also more generally [105].

The main mathematical mechanisms behind Theorem 2.2.2 have been introduced in the important contribution [129] for 'direct' statistical inference problems, see also [59]. The present form of this theorem relevant for non-linear inverse problems via 'rescaled' Gaussian priors that ensure sufficient 'posterior regularity' to permit the application of stability estimates is due to [91]. That reference [91] also contains the first proofs of results such as Theorems 2.3.1 and 2.3.2 in the non-linear setting. In the context of Darcy's problem, the consistency Theorem 2.3.3 was first obtained in [63] by adapting the ideas from [91]. For specific ('sieved') Gaussian process priors one can give improved results for the posterior regularity sets Θ_N from (2.24), see Ex. 2.4.3 and also [23, 102].

Regarding optimality of the results, one can prove that the rate δ_N obtained in (2.33) for the 'PDE-constrained regression problem' is optimal in a minimax sense (see [101], Theorem 10), which implies in particular that Proposition 2.1.3 cannot hold for $\kappa > 1$. In contrast the rate δ_N^η from (2.34) is unlikely to be sharp, since we have not attempted to optimise the exponent η in the stability estimate in Proposition 2.1.7. For 'smooth' conductivities where $\alpha, \beta \to \infty$, the rate in (2.34) *does* approach the optimal convergence rate $1/\sqrt{N}$ of finite-dimensional models. But the optimal reconstruction rate for Darcy's problem with general Sobolev regularity remains open at present. For the Schrödinger equation, the minimax optimal rate for inferring f in L^2-distance is $N^{-\alpha/(2\alpha+4+d)}$, see [94] and Exercises 2.4.1, 2.4.3. Our focus here is not on 'optimal' rates but to provide sufficient conditions for posterior consistency with 'algebraic' rates $N^{-a}, a > 0$, for a flexible class of commonly used Gaussian process priors. Such contraction rate results constitute the key 'localisation' step in the analysis of posterior measures in subsequent chapters.

Chapter 3

Information operators and curvature

The contraction rate theorems from Section 2.3 provide 'global' conditions on forward maps $\mathscr{G}$ that ensure that posterior distributions $\Pi(\cdot|D_N)$ (and their means) arising from Gaussian process priors concentrate on L^2_λ-balls centred at the ground truth θ_0. The radius of these balls was shown to shrink at rate N^{-a} for some $a > 0$, with N being sample size, and a depending on the Sobolev regularity of the prior Π, on analytical properties of the map $\mathscr{G}$, and on θ_0.

Now suppose one 'zooms in' to such a local neighbourhood of θ_0 of the posterior measure, and that $\mathscr{G}$ admits a good (say quadratic) approximation by its linearisation operator (derivative) $D\mathscr{G}_{\theta_0}$ at θ_0. In this case the posterior may start to display some remarkable universality features which are the subject of the remainder of these lecture notes. We will see that these features depend on the inverse problem only via $D\mathscr{G}_{\theta_0}$ and on an appropriate adjoint operator $D\mathscr{G}^*_{\theta_0}$, both to be defined carefully below. In analogy to the notion of the 'information matrix' used in classical (semi-) parametric statistics [127], one may call $D\mathscr{G}^*_{\theta_0} D\mathscr{G}_{\theta_0}$ the *information operator*. Both the information operator and $D\mathscr{G}^*_{\theta_0}$ encode a degree of 'local identifiability' of the statistical model near θ_0, and also quantify the 'local curvature' at θ_0 of the 'asymptotic least squares fit functional'. In this chapter we will review and lay out some of these ideas from semi-parametric statistics and then connect them to the high-dimensional inverse problems/PDE settings relevant here.

3.1 Information geometry for Gaussian non-linear regression models

The idea to describe 'information theoretic' features of statistical models $\{P_\theta : \theta \in \Theta\}$ by the derivative of the log-likelihood $(d/d\theta) \log dP_\theta$ at the 'true value' θ_0 is at least hundred years old and was advocated perhaps most prominently by

R.A. Fisher. See [127] for an account of the classical theory, whose main ideas and terminology we will follow here. In our non-linear regression model (1.12) these derivatives depend on the inverse problem via the linearisation of the forward map $\mathscr{G} : \Theta \to L^2_\lambda(\mathcal{X}, V)$ from (1.10), where we recall that the domain Θ is assumed to be a (measurable) subset of $L^2_\zeta(\mathcal{Z}, W)$.

The following condition stipulates the existence of a continuous derivative $D\mathscr{G}_{\theta_0}$ of $\mathscr{G}$ at θ_0 in given directions $h \in H$ where H is a linear *tangent space* of admissible directions. When the parameter space Θ is a linear space (as is relevant when using Gaussian process priors) it is natural to choose $\Theta = H$, but see after Definition 3.1.2 for more discussion.

Condition 3.1.1. *Let $\theta_0 \in \Theta$ and let $H \subset L^\infty(\mathcal{Z}, W)$ be a linear space such that the paths $\{\theta_0 + h, \|h\|_\infty < \epsilon, h \in H\}$, lie in Θ for some $\epsilon > 0$. Suppose the map $\mathscr{G}$ from (1.10) satisfies as $\|h\|_\infty \to 0$,*

$$\|\mathscr{G}(\theta_0 + h) - \mathscr{G}(\theta_0) - D\mathscr{G}_{\theta_0}[h]\|_{L^2_\lambda(\mathcal{X},V)} \equiv \rho_{\theta_0}[h] = o(\|h\|_\infty) \tag{3.1}$$

for some continuous linear operator

$$\mathbb{I}_{\theta_0} \equiv D\mathscr{G}_{\theta_0} : (H, \langle \cdot, \cdot \rangle_{L^2_\zeta(\mathcal{Z},W)}) \to L^2_\lambda(\mathcal{X}, V). \tag{3.2}$$

Variations of the previous condition can be considered too but the current form will be convenient for the PDE settings we have in mind.

By the Riesz-representation theorem, the operator $\mathbb{I}_{\theta_0}$ from (3.2) has a Hilbert space adjoint operator

$$\mathbb{I}^*_{\theta_0} : L^2_\lambda(\mathcal{X}, V) \to \overline{(H, \langle \cdot, \cdot \rangle_{L^2_\zeta(\mathcal{Z},W)})} \equiv \bar{H}, \tag{3.3}$$

(with $\overline{H}$ denoting Hilbert-space completion), for which

$$\langle \mathbb{I}_{\theta_0} h, g \rangle_{L^2_\lambda} = \langle h, \mathbb{I}^*_{\theta_0} g \rangle_{L^2_\zeta}, \quad h \in H, g \in L^2_\lambda.$$

This then leads to the following definition:

Definition 3.1.2. *For a forward map $\mathscr{G}$ satisfying Condition 3.1.1 for tangent space H and operator $\mathbb{I}_\theta$ from (3.2), the information operator is defined as*

$$\mathbb{I}^*_{\theta_0}\mathbb{I}_{\theta_0} : (H, \langle \cdot, \cdot \rangle_{L^2_\zeta(\mathcal{Z},W)}) \to \overline{(H, \langle \cdot, \cdot \rangle_{L^2_\zeta(\mathcal{Z},W)})}.$$

Of course we can extend $\mathbb{I}^*_{\theta_0}\mathbb{I}_{\theta_0}$ to all of $\overline{H}$ by continuity. We note that at this stage, the choice of Θ is not just one of a natural domain for the map $\mathscr{G}$, but also represents a commitment on the statistical model as it determines the (maximal) tangent space H. For instance if $\Theta = H$, the 'complexity' of the set H of paths near θ_0 is determined by the complexity of the parameter space Θ. In this case, restricting a priori to a L^2_ζ-closed subspace $\Theta_0 \subset \Theta$ changes the adjoint operator $\mathbb{I}^*_{\theta_0}$ and then also the information operator. But in our infinite-dimensional setting, L^2_ζ-*dense* subspaces give rise to the same definition – for example choosing $\Theta = H^\beta(\mathcal{Z})$ or $\Theta = L^\infty(\mathcal{Z})$ makes no difference as their completions for the L^2_ζ-norms are both L^2_ζ (when $\mathcal{Z}$ is a bounded smooth domain in $\mathbb{R}^d$ and ζ Lebesgue measure, say).

3.1.1 The LAN expansion

A large class of 'regular' statistical models $\{P_\theta : \theta \in \Theta\}$ admit what is called a 'locally asymptotically normal' (LAN) expansion of the logarithms of the likelihood ratios

$$\log \frac{dP_{\theta_0+h/\sqrt{N}}}{dP_{\theta_0}}(D_N) \approx N\Big(-\frac{1}{2}\|h\|^2_{LAN}, \|h\|^2_{LAN}\Big), \quad \theta_0 \in \Theta, \tag{3.4}$$

along local paths $\theta_0 + h/\sqrt{N} \in \Theta, h \in H$, assuming data $D_N \sim P^N_{\theta_0}$, and where $\langle \cdot, \cdot \rangle_{LAN}$ defines an inner product on the tangent space H with 'LAN-norm' $\|\cdot\|_{LAN}$. The distribution on the r.h.s. is not just normal but perhaps more importantly also coincides with the distribution of the log-likelihood ratio (Radon-Nikodym density)

$$\log \frac{dP_h}{dP_0}(Z), \ Z \sim dP_0, \tag{3.5}$$

where dP_h is the distribution of a Gaussian variable $Z = h + \mathbb{W}$ with unknown mean 'shift' parameter $h \in H$ and $\mathbb{W}$ a centred Gaussian white noise process on $(H, \langle \cdot, \cdot \rangle_{LAN})$. [We refer to [61] for the definition of such processes as well as to the Cameron-Martin theorem required to show that the distribution of (3.5) indeed equals the r.h.s. on (3.4).] The gist of Le Cam theory [83, 127, 130] is that in LAN settings the asymptotic inference problem in the model $\{P_\theta : \theta \in \Theta\}$ can locally near θ_0 be regarded as the much simpler one of inferring the shift h from a Gaussian observation $Z \sim dP_h$.

For the Gaussian regression model (1.11) such a LAN expansion holds as soon as the forward map $\mathscr{G}$ satisfies some natural conditions.

Theorem 3.1.3. *Suppose that Condition 3.1.1 holds for some $\theta_0 \in \Theta$ and fix any $h \in H \subset L^\infty(\mathcal{Z})$. Assume further that for some constant $C = C_{\mathscr{G},h,\epsilon}$,*

$$\|\mathscr{G}(\theta_0 + sh) - \mathscr{G}(\theta_0)\|_\infty \leq C|s|, \quad \textit{for all } |s| \textit{ small enough.} \tag{3.6}$$

Then the log-likelihood ratio process in the model (1.11) with marginal densities given in (1.12) satisfies, as $N \to \infty$, the asymptotic expansion

$$\log \frac{dP^N_{\theta_0+h/\sqrt{N}}}{dP^N_{\theta_0}}(D_N) = W_N(h) - \frac{1}{2}\|\mathbb{I}_{\theta_0}[h]\|^2_{L^2_\lambda(\mathcal{X},V)} + o_{P^N_{\theta_0}}(1) \tag{3.7}$$

for random variables

$$W_N \equiv -\frac{1}{\sqrt{N}} \sum_{i=1}^N \langle \mathbb{I}_{\theta_0}[h](X_i), \varepsilon_i \rangle_V \to^d N(0, \|\mathbb{I}_{\theta_0}[h]\|^2_{L^2_\lambda(\mathcal{X},V)}), \tag{3.8}$$

where $\to^d$ *denotes convergence in distribution (under* $P^{\mathbb{N}}_{\theta_0}$*) of real random variables. In particular, the model satisfies the LAN-approximation (3.4): as* $N \to \infty$,

$$\log \frac{dP^N_{\theta_0+h/\sqrt{N}}}{dP^N_{\theta_0}}(D_N) \to^d N\Big(-\frac{1}{2}\|\mathbb{I}_{\theta_0}[h]\|^2_{L^2_\lambda(\mathcal{X},V)}, \|\mathbb{I}_{\theta_0}[h]\|^2_{L^2_\lambda(\mathcal{X},V)}\Big). \tag{3.9}$$

Proof. We recall the notation (1.15) so that $\ell_N(\theta)$ equals $\log dP^N_\theta$ up to universal additive constants. Then under $P^N_{\theta_0}$ we have

$$\begin{aligned}
&\ell_N(\theta_0) - \ell_N(\theta_0 + h/\sqrt{N}) \\
&= -\frac{1}{2}\sum_{i=1}^N \big(|Y_i - \mathscr{G}(\theta_0)(X_i)|_V^2 - |Y_i - \mathscr{G}(\theta_0 + h/\sqrt{N})(X_i)|_V^2\big) \\
&= -\frac{1}{2}\sum_{i=1}^N \big(|\varepsilon_i|_V^2 - |\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta_0 + h/\sqrt{N})(X_i) + \varepsilon_i|_V^2\big) \\
&= \sum_{i=1}^N \langle \varepsilon_i, \mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta_0 + h/\sqrt{N})(X_i)\rangle_V \\
&\qquad + \frac{1}{2}\sum_{i=1}^N |\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta_0 + h/\sqrt{N})(X_i)|_V^2\big) \equiv I + II.
\end{aligned}$$

We can write

$$\begin{aligned}
I = &\frac{1}{\sqrt{N}}\sum_{i=1}^N \langle \varepsilon_i, D\mathscr{G}_{\theta_0}[h](X_i)\rangle_V \\
&+ \sum_{i=1}^N \langle \varepsilon_i, \mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta_0 + h/\sqrt{N})(X_i) - D\mathscr{G}_{\theta_0}[h/\sqrt{N}](X_i)\rangle_V.
\end{aligned}$$

The first summand amounts to the term W_N in (3.7) and (3.8) then follows from $\mathbb{I}_{\theta_0}[h] \in L^2_\lambda(\mathcal{X}, V)$ and the central limit theorem. By independence and Condition 3.1.1 the last summand has variance bounded by the second moment

$$NE_\lambda\big|\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta_0 + h/\sqrt{N})(X_i) - D\mathscr{G}_{\theta_0}[h/\sqrt{N}](X_i)\big|_V^2 = N\rho^2_{\theta_0}[h/\sqrt{N}] = o(1),$$

hence by Markov's inequality is $o_{P^N_{\theta_0}}(1)$ and negligible in the asymptotic distribution. For term II, we first center at its E_λ expectation

$$\frac{1}{2}\sum_{i=1}^N \Big(|\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta_0 + h/\sqrt{N})(X_i)|_V^2 - E_\lambda|\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta_0 + h/\sqrt{N})(X_i)|_V^2\Big)$$
$$+ \frac{N}{2}\|\mathscr{G}(\theta_0) - \mathscr{G}(\theta_0 + h/\sqrt{N})\|^2_{L^2(\mathcal{X})}. \tag{3.10}$$

For the first term we can bound the variance of the summands

$$Z_i = |\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta_0 + h/\sqrt{N})(X_i)|_V^2$$

via the local Lipschitz property (3.6) as

$$EZ_i^2 \le \|\mathscr{G}(\theta_0) - \mathscr{G}(\theta_0 + h/\sqrt{N})\|_\infty^4 \lesssim \frac{1}{N^2}$$

so that by independence and Chebyshev's inequality $Var(\sum_{i=1}^N Z_i/2) \lesssim NN^{-2} \to 0$ for fixed h and hence $\sum_i Z_i/2 = o_{P^N_{\theta_0}}(1)$ is asymptotically negligible. We finally note that

$$\begin{aligned}
&\|\mathscr{G}(\theta_0 + h/\sqrt{N}) - \mathscr{G}(\theta_0) - D\mathscr{G}_{\theta_0}[h/\sqrt{N}] + D\mathscr{G}_{\theta_0}[h/\sqrt{N}]\|^2_{L^2(\mathcal{X})}\\
&= \|D\mathscr{G}_{\theta_0}[h/\sqrt{N}]\|^2_{L^2(\mathcal{X})}\\
&\quad + \frac{2}{\sqrt{N}}\langle D\mathscr{G}_{\theta_0}[h], \mathscr{G}(\theta_0 + h/\sqrt{N}) - \mathscr{G}(\theta_0) - D\mathscr{G}_{\theta_0}[h/\sqrt{N}]\rangle_{L^2(\mathcal{X})}\\
&\quad + \|\mathscr{G}(\theta + h/\sqrt{N}) - \mathscr{G}(\theta_0) - D\mathscr{G}_{\theta_0}[h/\sqrt{N}]\|^2_{L^2(\mathcal{X})}\\
&= \frac{1}{N}\|D\mathscr{G}_{\theta_0}[h]\|^2_{L^2(\mathcal{X})} + o(1/N)
\end{aligned}$$

where we have used Condition 3.1.1. Therefore the second term in (3.10) features the second term required in the r.h.s. of (3.7) plus asymptotically negligible terms, completing the proof of (3.9). □

3.1.2 Cramer-Rao bounds and inverse information

We can analyse finer properties of inference problems with infinite-dimensional parameter spaces Θ by studying linear functionals $\langle\theta, \psi\rangle_{L^2_\zeta}, \theta \in \Theta$, where ψ runs through a collection of 'test functions'. Following standard ideas from functional analysis we can 'learn' various properties of an infinite-dimensional problem by understanding sufficiently large collections of such one-dimensional subproblems. In this section we use ideas from semi-parametric statistics to derive what 'optimal' procedures can attain in the LAN model from Theorem 3.1.3 in dependence of analytical properties of the test function ψ and the operator $\mathbb{I}_{\theta_0}$. These ideas set the stage for studying universal Gaussian approximations to Bayesian posterior measures in Chapter 4.

The local asymptotic minimax bound for linear functionals

For H the tangent space from Condition 3.1.1, suppose for now that a linear functional $\Psi : \Theta \to \mathbb{R}$ can be represented as

$$\Psi(h) = \langle \tilde{\psi}_{\theta_0}, \mathbb{I}_{\theta_0} h \rangle_{L^2_\lambda}, \quad h \in H; \quad \text{for some } \tilde{\psi}_{\theta_0} \in L^2_\lambda(\mathcal{X}, V). \tag{3.11}$$

Notice that by the Cauchy-Schwarz inequality such Ψ is continuous on H for the LAN-norm $\|\mathbb{I}_{\theta_0}(\cdot)\|_{L^2_\lambda}$. Whenever such $\tilde{\psi}_{\theta_0}$ exists we can take it to belong to the closure $\overline{\mathscr{I}_H}$ of the linear space

$$\mathscr{I}_H = \{w = \mathbb{I}_{\theta_0}(h) : h \in H\}$$

in L^2_λ, by orthogonal projection onto $\overline{\mathscr{I}_H}$ if necessary. Assuming (3.11) and the LAN expansion from Theorem 3.1.3, the following is a classical result from mathematical statistics [127, 130]. We assume that $\mathbb{I}_{\theta_0}$ is injective but this is not necessary (see Ex. 3.4.1).

Theorem 3.1.4. *Consider data $(Y_i, X_i)_{i=1}^N$ in the model (1.11) where $\mathscr{G}$ satisfies the hypotheses of Theorem 3.1.3 with tangent space H. Suppose that $\mathbb{I}_{\theta_0} : H \to L^2_\lambda$ is injective and that the functional $\Psi : \Theta \to \mathbb{R}$ satisfies (3.11). Then the local asymptotic minimax risk for estimating Ψ at θ_0 is lower bounded as*

$$\liminf_{N\to\infty} \inf_{\bar{\psi}_N : (V\times\mathcal{X})^N \to \mathbb{R}} \sup_{h\in H : \|h\|_{L^2_\zeta} \le 1/\sqrt{N}} N E^N_{\theta_0+h}(\bar{\psi}_N - \Psi(\theta_0 + h))^2 \geq \|\tilde{\psi}_{\theta_0}\|^2_{L^2_\lambda}. \tag{3.12}$$

Proof. The result follows from Theorem 3.11.5 in [130] applied to the LAN model from Theorem 3.1.3. In the former theorem we choose quadratic loss $\ell = (\cdot)^2$, Banach space $\mathbf{B} = \mathbb{R}$ (so existence of a tight Gaussian limit is implied by finiteness of $\|\tilde{\psi}_{\theta_0}\|_{L^2_\lambda}$) and with Hilbert space norm $\|\cdot\| = \|\mathbb{I}_{\theta_0}(\cdot)\|_{L^2_\lambda}$ on the tangent space H, noting also that the map Ψ satisfies, by (3.11) and the Cauchy-Schwarz inequality,

$$|\Psi(h)| \le \|\tilde{\psi}_{\theta_0}\|_{L^2_\lambda} \|h\|$$

and is hence continuous from $H \to \mathbf{B}$. $\square$

In particular for functionals Ψ as in (3.11) there is hope for the existence of efficient $\sqrt{N}$-consistent estimators that attain this lower bound. Let us shed some more light on the condition (3.11). If Ψ is of the more natural form

$$\Psi(\theta) = \langle \theta, \psi \rangle_{L^2_\zeta}, \quad \text{for some } \psi \in L^2_\zeta, \tag{3.13}$$

then (3.11) can be rewritten as requiring that

$$\psi = \mathbb{I}_{\theta_0}^* \tilde{\psi}_{\theta_0} \text{ for some } \tilde{\psi}_{\theta_0} \in L^2_\lambda, \tag{3.14}$$

in other words, that ψ lies in the range $R(\mathbb{I}_\theta^*)$ of the adjoint operator $\mathbb{I}_{\theta_0}^*$. This is related to the local identifiability of our model via the standard Hilbert space identity ($\perp$ denoting orthogonal complement)

$$ker(\mathbb{I}_{\theta_0}) = R(\mathbb{I}_{\theta_0}^*)^\perp, \tag{3.15}$$

so requires in particular that ψ does not lie in the kernel of $\mathbb{I}_{\theta_0}$ (acting on L^2_ζ). But note that unless the range of $\mathbb{I}_{\theta_0}^*$ is closed, even if $\psi \notin ker(\mathbb{I}_{\theta_0})$ so that $\psi \in \overline{R(\mathbb{I}_{\theta_0}^*)}$, it need not belong to $R(\mathbb{I}_{\theta_0}^*)$.

A sufficient (but not necessary) condition for (3.14) is that ψ lies in the range of the information operator $\mathbb{I}_{\theta_0}^* \mathbb{I}_{\theta_0}$. Indeed if $\bar{\psi} \in H$ solves the *information equation*

$$\mathbb{I}_{\theta_0}^* \mathbb{I}_{\theta_0} \bar{\psi} = \psi \tag{3.16}$$

for the given ψ, then we can use $\tilde{\psi}_{\theta_0} = \mathbb{I}_{\theta_0} \bar{\psi} \in L^2_\lambda$ in (3.11). We will turn to this issue again in Section 4.1.

The information number and its inverse

When (3.11) holds the lower bound in Theorem 3.1.4 has variational characterisation

$$\|\tilde{\psi}_{\theta_0}\|^2_{L^2_\lambda} = \sup_{0 \neq w = \mathbb{I}_{\theta_0}(h), h \in H} \frac{\langle \tilde{\psi}_{\theta_0}, w \rangle^2_{L^2_\lambda}}{\langle w, w \rangle_{L^2_\lambda}}. \tag{3.17}$$

Indeed, the Cauchy-Schwarz inequality gives one side of the inequality and the converse follows from taking a sequence $w \in \mathscr{I}_H$ approaching $\tilde{\psi}_{\theta_0} \in \overline{\mathscr{I}_H}$.

When the target of inference is a functional of the form $\Psi(\theta) = \langle \theta, \psi \rangle_{L^2}$ from (3.13), then the characterisation (3.17) can be related to the *efficient information number* $i_{\theta_0, H, \psi}$ for (unbiased) estimation of Ψ over the sub-models $\{\theta_0 + sh : |s| < \epsilon, h \in H\}$ appearing in the classical Cramér-Rao theorem: We first set

$$i_{\theta_0, h, \psi} := \frac{\|\mathbb{I}_{\theta_0} h\|^2_{L^2_\lambda}}{\langle \psi, h \rangle^2_{L^2_\zeta}}, \quad h \in H,$$

and define

$$i_{\theta_0, H, \psi} \equiv \inf_{h \in H, \langle \psi, h \rangle_{L^2_\zeta} \neq 0} \frac{\|\mathbb{I}_{\theta_0} h\|^2_{L^2_\lambda}}{\langle \psi, h \rangle^2_{L^2_\zeta}}. \tag{3.18}$$

If the range condition (3.14) holds then we can rewrite the denominator as $\langle \mathbb{I}_{\theta_0}^* \tilde{\psi}_{\theta_0}, h\rangle_{L^2_\zeta}^2 = \langle \tilde{\psi}_{\theta_0}, \mathbb{I}_{\theta_0} h\rangle_{L^2_\lambda}^2$. Also (3.15) implies that ψ is then necessarily orthogonal on the kernel of $\mathbb{I}_{\theta_0}$ and we deduce

$$\|\tilde{\psi}_{\theta_0}\|_{L^2_\lambda}^2 = \sup_{h\in H:\mathbb{I}_{\theta_0}h\neq 0} \frac{\langle \tilde{\psi}_{\theta_0}, \mathbb{I}_{\theta_0} h\rangle_{L^2_\lambda}^2}{\langle \mathbb{I}_{\theta_0} h, \mathbb{I}_{\theta_0} h\rangle_{L^2_\lambda}} = i_{\theta_0,H,\psi}^{-1}. \tag{3.19}$$

Non-existence of the inverse information

When the information number $i_{\theta_0,H,\psi}$ is non-zero so that the quantity (3.19) is finite, the lower bound from Theorem 3.1.4 can often be 'matched' by non-parametric Bayesian methods by i) solving the information equation (3.16) and by ii) making the techniques underlying the proof of Theorem 3.1.3 quantitative and 'uniform' in relevant classes of alternatives h – this will be the content of Section 4.1 below.

In contrast, if the lower bound in Theorem 3.1.4 is *infinite*, this introduces fundamental limitations to the proof methods arising from LAN expansions. Specifically, when considering inference on functionals Ψ from (3.13), the following theorem shows that the adjoint range condition (3.14) is necessary for the existence of a non-zero information number (3.18), and for the possibility of locally uniformly $\sqrt{N}$-consistent estimation of the scalar parameter $\Psi(\theta)$. This will become relevant in Chapter 4, too.

Theorem 3.1.5. *Consider data $(Y_i, X_i)_{i=1}^N$ in the model (1.11) where $\mathscr{G}$ satisfies the hypotheses of Theorem 3.1.3 with tangent space H and adjoint score operator $\mathbb{I}_{\theta_0}^*$ from (3.3). Consider a functional $\Psi(\theta) = \langle \theta, \psi\rangle_{L^2_\zeta}$ for some $\psi \in L^2_\zeta$ for which*

$$\psi \notin R(\mathbb{I}_{\theta_0}^*) := \{h = \mathbb{I}_{\theta_0}^* g \text{ for some } g \in L^2_\lambda(\mathcal{X}, V)\}. \tag{3.20}$$

Then the efficient information number from (3.18) equals zero, $i_{\theta_0,H,\psi} = 0$, and

$$\liminf_{N\to\infty} \inf_{\bar{\psi}_N:(V\times\mathcal{X})^N\to\mathbb{R}} \sup_{h\in H:\|h\|_{L^2_\zeta}\leq 1/\sqrt{N}} N E_{\theta_0+h}^N (\bar{\psi}_N - \Psi(\theta_0 + h))^2 = \infty. \tag{3.21}$$

Proof. That $i_{\theta_0,H,\psi} = 0$ follows from a functional analytic argument showing that the ratio in (3.18) necessarily vanishes if (3.20) holds – see Theorem 6.2.8 in the appendix for the proof.

Then assuming $i_{\theta_0,H,\psi} = 0$, we augment the observation space to include measurements $(Z_i, Y_i, X_i)_{i=1}^N \sim \bar{P}_{\theta_0}^N$ where the

$$Z_i \sim^{iid} N(\langle \theta_0, \psi\rangle_{L^2_\zeta}, \sigma^2)$$

are independent of the (Y_i, X_i)'s, and where $\sigma^2 > 0$ is known but arbitrary. The new statistical model $\{\bar{P}_\theta^N : \theta \in \Theta\}$ again verifies the LAN approximation (3.4) with 'augmented' LAN norm

$$\|h\|_{LAN}^2 = \|\bar{\mathbb{I}}_{\theta_0} h\|_{L^2_\lambda}^2 = \|\mathbb{I}_{\theta_0} h\|_{L^2_\lambda}^2 + \sigma^{-2}\langle \psi, h\rangle_{L^2_\zeta}^2, \quad h \in H, \tag{3.22}$$

as follows from Theorem 3.1.3, a standard tensorisation argument for the LAN property in independent sample spaces, and the fact that a $\mathcal{N}(\langle \theta_0, \psi\rangle_{L^2}, \sigma^2)$ model has LAN 'norm' $\sigma^{-2}\langle \psi, h\rangle_{L^2}^2$, by a simple direct calculation with Gaussian densities. The new efficient information from (3.18) for estimating $\Psi(\theta)$ from the augmented data (and with identical tangent space) is now of the form

$$\bar{i}_{\theta_0,H,\psi} = \inf_h \frac{\|\mathbb{I}_{\theta_0} h\|_{L^2_\lambda}^2 + \sigma^{-2}\langle \psi, h\rangle_{L^2_\zeta}^2}{\langle \psi, h\rangle_{L^2_\zeta}^2} = i_{\theta_0,H,\psi} + \sigma^{-2} = \sigma^{-2} > 0$$

where we have used the fact that $i_{\theta_0,H,\psi} = 0$.

To proceed we note that the linear functional $\Psi(\cdot)$ is continuous on H for the augmented LAN-norm (3.22) so that we can invoke the Riesz representation theorem on $\bar{H}$ (or, if $\bar{\mathbb{I}}_{\theta_0}$ is not injective on $\bar{H}$, on the quotient space of $\bar{H}$ for the kernel of $\bar{\mathbb{I}}_{\theta_0}$) to the effect that

$$\Psi(h) = \langle \bar{\mathbb{I}}_{\theta_0}\tilde{h}, \bar{\mathbb{I}}_{\theta_0} h\rangle_{L^2_\lambda}, \quad h \in H, \text{ for some } \tilde{\psi}_{\theta_0} \equiv \bar{\mathbb{I}}_{\theta_0}\tilde{h} \in \overline{(\mathscr{I}_H)}.$$

In particular Ψ now verifies (3.11) in the augmented settting and one then shows (see Ex. 3.4.2) that mutatis mutandis, (3.12) and (3.19) hold in the augmented model as well, with linearisation operator $\bar{\mathbb{I}}_{\theta_0}$ and tangent space H. In particular

$$\liminf_{N\to\infty} \inf_{\bar{\psi}_N : (\mathbb{R}\times V\times\mathcal{X})^N \to \mathbb{R}} \sup_{h\in H, \|h\|_{L^2_\zeta} \le 1/\sqrt{N}} N E_{\theta_0+h}^N (\bar{\psi}_N - \Psi(\theta_0 + h))^2 \ge \bar{i}_{\theta_0,H,\psi}^{-1} = \sigma^2 \tag{3.23}$$

for estimators $\bar{\psi}$ based on the more informative data. The asymptotic local minimax risk in (3.21) exceeds the quantity in the last display, and letting $\sigma^2 \to \infty$ implies the result. □

3.2 Gradient stability, local convexity, and concentration

In this section we look at another aspect that is determined by the linearisation operator $D\mathscr{G}_{\theta_0} = \mathbb{I}_{\theta_0}$ of $\mathscr{G}$ at the ground truth θ_0 – namely the *local curvature* of the log-likelihood function $\ell_N(\theta)$ from (1.15) featuring in the expression of the posterior measure (1.14). We will show that under certain conditions on $D\mathscr{G}_{\theta_0}$, the 'least squares fit' $-\ell_N(\theta)$ is locally convex near the ground truth θ_0 with high $P_{\theta_0}^N$-probability. This implies that the posterior measure $\Pi(\cdot|D_N)$ arising from Gaussian priors will be locally *log-concave* in that region, which will be exploited further in Chapter 5.

3.2.1 Convexity of $-\ell_N$ and the gradient of $\mathscr{G}$

We introduce *finite-dimensional* spaces

$$\mathbb{R}^D \simeq span\{e_j : 1 \le j \le D\} \equiv E_D \subset \Theta \tag{3.24}$$

arising from a set of orthonormal basis functions $e_j \in L^2_\zeta(\mathcal{Z}, W)$ used to 'discretise' the ambient linear parameter space $\Theta \subset L^2_\zeta$. We denote the Euclidean norm on $E_D \simeq \mathbb{R}^D$ by $\|\cdot\|_{E_D}$. A prototypical example pursued below is to take the eigenfunctions of the Dirichlet-Laplacian Δ on a bounded smooth domain $\mathcal{Z}$ as in Ex. 2.4.3 from earlier. While D is finite, we think of it as 'high-dimensional' since letting $D \to \infty$ will reduce the approximation error. The gradient vectors $\nabla\mathscr{G}(\theta)$ of $\mathscr{G}(\cdot)(x) : E_D \to \mathbb{R}, x \in \mathcal{X}$, are now defined via partial derivatives

$$\frac{\partial}{\partial \theta_j}\mathscr{G}(\theta) = \frac{\partial}{\partial t_j}\mathscr{G}\Big(\sum_{j=1}^D t_j e_j\Big), \quad t_j = \langle \theta, e_j\rangle_{L^2_\zeta}, \tag{3.25}$$

where $\theta = \sum_{j=1}^D t_j e_j \in E_D$.

In view of the non-linearity of $\mathscr{G}$, the 'least squares fit' $-\ell_N(\theta) = -\sum_{i=1}^N \ell_i(\theta)$ from (1.15) is neither globally nor locally convex for the given observation vector. Indeed, assuming $\mathscr{G}$ is sufficiently smooth, the empirical Hessian of $-\ell_N(\theta)$ equals the sum of all

$$-\nabla^2\ell_i(\theta) = [\nabla\mathscr{G}(\theta)(X_i)][\nabla\mathscr{G}(\theta)(X_i)]^T + [\mathscr{G}(\theta)(X_i) - Y_i][\nabla^2\mathscr{G}(\theta)(X_i)]. \tag{3.26}$$

Quadratic forms of the last expression may have an arbitrary sign, different for distinct θ, and so the Hessian matrix is not necessarily positive semi-definite. But if we assume for the moment that $\theta_0 \in E_D$ (later we will consider its projection $\theta_{0,D}$ onto E_D) and take E_{θ_0} expectations under the model equation (1.11) in the last identity, we see that the second term in (3.26) vanishes at $\theta = \theta_0$, and that hence for any $v \in E_D \simeq \mathbb{R}^D$,

$$v^T[-E_{\theta_0}\nabla^2\ell_i(\theta_0)]v = \|\nabla\mathscr{G}(\theta_0)^T v\|^2_{L^2_\lambda}. \tag{3.27}$$

We deduce that if the gradient $\nabla\mathscr{G}(\theta)$ satisfies a 'gradient stability' estimate

$$\|\nabla\mathscr{G}(\theta_0)^T v\|^2_{L^2_\lambda} \gtrsim \|v\|^2_{E_D}, \tag{3.28}$$

then *on average* the negative log-likelihood function has a positive definite Hessian at θ_0, and we expect this bound to extend (by continuity) to $-E_{\theta_0}\nabla^2\ell(\theta)$ at least for θ in a neighbourhood of θ_0. We can then use tools from high-dimensional probability to deduce that the discretely sampled functions ℓ_N concentrate around their statistical means sufficiently well to inherit the local curvature.

The hypothesis (3.28) is a quantitative injectivity condition on the linearisation operator $D\mathscr{G}_{\theta_0}$ from Condition 3.1.1 expressed in terms of its action on the discretisation space E_D. Under Condition 3.1.1 with $H \supset E_D$, the requirement (3.28) becomes

$$\langle \nabla\mathscr{G}(\theta_0)^T v, \nabla\mathscr{G}(\theta_0)^T v\rangle_{L^2_\lambda} = \langle D\mathscr{G}_{\theta_0}[v], D\mathscr{G}_{\theta_0}[v]\rangle_{L^2_\lambda} = \langle v, D\mathscr{G}^*_{\theta_0} D\mathscr{G}_{\theta_0}[v]\rangle_{L^2_\zeta}, \quad (3.29)$$

and hence amounts to a lower bound on the minimal eigenvalue of the matrix $\nabla\mathscr{G}(\theta_0)\nabla\mathscr{G}(\theta_0)^T$ arising from the L^2_ζ-action of the information operator $\mathbb{I}^*_{\theta_0}\mathbb{I}_{\theta_0}$ from (3.1.2) on the finite dimensional subspace E_D. Such a result is implied in particular by an infinite-dimensional stability estimate

$$\|D\mathscr{G}_{\theta_0}[h]\|_{L^2_\lambda} = \|\mathbb{I}_{\theta_0}[h]\|_{L^2_\lambda} \gtrsim \|h\|_+, \quad h \in H, \quad (3.30)$$

for an appropriate $\|\cdot\|_+$ norm that is 'compatible' with the E_D scale. When E_D is the eigen-basis of the Dirichlet-Lapacian we will verify this condition for our model PDEs – see Theorem 5.3.2 for Darcy's problem and (3.60) as well as Ex. 5.4.5 for the Schrödinger equation.

3.2.2 A concentration result for the empirical Hessian

We will require 'local' conditions near a fixed $\theta_0 \in \Theta \subset L^2_\zeta(\mathcal{Z}, W)$, in fact near its L^2_ζ-orthonormal projection $\theta_{0,D}$ onto E_D. For radius $\mathbf{r} > 0$ to be chosen, set

$$\mathcal{B} := \big\{\theta \in E_D : \|\theta - \theta_{0,D}\|_{E_D} < \mathbf{r}\big\}, \quad (3.31)$$

where we write $\|\cdot\|_{E_D} \simeq \|\cdot\|_{\mathbb{R}^D}$ for the L^2_ζ-norm on E_D in this subsection.

The following local regularity condition on the map $\mathscr{G}$ is satisfied, for instance, as soon as $\mathscr{G}$ is C^3 on $\mathcal{B}$ with local C^3-norm constants growing at most *polynomially* in dimension D. To formulate it let us define the following local $C^{2,1}$-norm for maps $F : \mathcal{B} \to V$, $F(\theta) = (F_1(\theta), \dots, F_{d_V}(\theta))$,

$$\|F\|_{C^{2,1}(\mathcal{B},V)} := \max_{1\le k\le d_V} \Big\{\|F_k\|_\infty + \|\nabla F_k\|_{L^\infty(\mathcal{B},\mathbb{R}^D)} + \|\nabla^2 F_k\|_{L^\infty(\mathcal{B},\mathbb{R}^{D\times D})} \quad (3.32)$$
$$+ \sup_{\theta,\theta'\in\mathcal{B},\theta\neq\theta'} \frac{|\nabla^2 F_k(\theta) - \nabla^2 F_k(\theta')|_{\mathrm{op}}}{\|\theta-\theta'\|_{E_D}}\Big\},$$

where $\|\cdot\|_{L^\infty(\mathcal{B},U)}$ are the supremum norms of maps defined on $\mathcal{B}$ taking values in a normed vector space U, the space $\mathbb{R}^{D\times D}$ is equipped with the usual operator norm $|\cdot|_{\mathrm{op}}$, and $\nabla=\nabla_\theta$ and ∇^2 denote the gradient and 'Hessian' operator, respectively, arising via (3.25).

Condition 3.2.1 (Local regularity). *Let $\mathcal{B}$ be given in (3.31) for some $\mathbf{r}>0$ and suppose that for all $x\in\mathcal{X}$, the map $\theta\mapsto\mathscr{G}(\theta)(x)$ from (1.10) is in $C^{2,1}(\mathcal{B},V)$ and satisfies $\sup_{x\in\mathcal{X}}\left\|\mathscr{G}(\cdot)(x)\right\|_{C^{2,1}(\mathcal{B},V)}\leq c_2D^{\kappa_2}$ for some $c_2\geq 1$ and $\kappa_2\geq 0$.*

For the next condition let us define

$$\ell(\theta,(Y,X)):=-\frac{1}{2}|Y-\mathscr{G}(\theta)(X)|_V^2,\quad \theta\in\Theta, \tag{3.33}$$

which equals ℓ_N from (1.15) for a single 'generic' observation (Y,X) from model (1.11). Regarding $\ell(\cdot,(Y,X))$ as a real-valued map defined on E_D, Condition 3.2.1 and the chain rule imply that the gradient vector $\nabla\ell(\theta,(Y,X))$ and Hessian matrix $\nabla^2\ell(\theta,(Y,X))$ with respect to $\theta\in E_D$ exist for every (Y,X). In particular the $D\times D$ matrix $E_{\theta_0}[-\nabla^2\ell(\theta,(Y,X))]$ is symmetric and $\lambda_{min}(A)$ will denote the smallest eigenvalue of a symmetric matrix A.

Condition 3.2.2 (Local average curvature). *Let $\mathcal{B}$ be given in (3.31) for some $\mathbf{r}>0$. Assume that for some $c_0>0, c_1\geq 1$, $\kappa_0,\kappa_1\geq 0$ and all $D\in\mathbb{N}$,*

$$\inf_{\theta\in\mathcal{B}}\lambda_{min}\Big(E_{\theta_0}[-\nabla^2\ell(\theta,(Y,X)))]\Big)\geq c_0D^{-\kappa_0}\quad\text{and} \tag{3.34}$$

$$\sup_{\theta\in\mathcal{B}}\Big[|E_{\theta_0}\ell(\theta,(Y,X))|+\|E_{\theta_0}[\nabla\ell(\theta,(Y,X))]\|_{\mathbb{R}^D}+\|E_{\theta_0}[\nabla^2\ell(\theta,(Y,X))]\|_{\mathrm{op}}\Big]\leq c_1D^{\kappa_1}. \tag{3.35}$$

As discussed in the preceding subsection, verification of (3.34) for $\mathcal{B}$ a neighbourhood of $\theta_{0,D}$ can be reduced to the condition (3.28) combined with Lipschitz continuity of the Hessian with respect to the θ-variable. Condition (3.35) is implied by Condition 3.2.1 for $\kappa_1=\kappa_2$ but we still record it separately to permit a potentially better constant $\kappa_1<\kappa_2$.

The following theorem is the main result of this subsection. It is based on tools from high-dimensional probability. The constraint $\mathbf{r}\leq 1$ is not necessary but convenient to permit a simple expression in (3.36) for the concentration exponent $\mathcal{R}_N$, and also natural for the applications we have in mind. A more detailed version of this result can be found in [102].

Theorem 3.2.3. *Let $\ell_N:E_D\to\mathbb{R}$ be given by (1.15). Suppose Conditions 3.2.1, 3.2.2 are satisfied for $\mathcal{B}$ from (3.31) and some $\mathbf{r}\leq 1$. There exists a constant $C=C(c_0,c_1,c_2)>0$ such that if*

$$\mathcal{R}_N:=CND^{-2\kappa_0-4\kappa_2}, \tag{3.36}$$

then for any $D, N \geq 1$ *satisfying* $D \leq \mathcal{R}_N$*, we have for constants* c, c' *depending only on* d_V

$$P_{\theta_0}^N\Big(\inf_{\theta\in\mathcal{B}} \lambda_{min}\big[-\nabla^2\ell_N(\theta)\big] < \frac{c_0}{2}ND^{-\kappa_0}\Big) \leq ce^{-\mathcal{R}_N}, \quad \textit{as well as} \tag{3.37}$$

$$P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\Big[|\ell_N(\theta)|+\|\nabla\ell_N(\theta)\|_{\mathbb{R}^D}+\|\nabla^2\ell_N(\theta)\|_{\mathrm{op}}\Big] > N(5c_1D^{\kappa_1}+1)\Big) \leq c'(e^{-\mathcal{R}_N}+e^{-N/8}). \tag{3.38}$$

Proof. It suffices to prove the assertion for $\mathcal{R}_N \geq 1$. We also restrict to the case $V = W = \mathbb{R}$ relevant in these notes – considering general vector-spaces V, W of fixed finite dimension is only notationally different (e.g., Lemma 5.6 in [23]).

For the proof we require some more notation: For $x \in \mathcal{X}$, we write shorthand $\mathscr{G}^x(\theta) := \mathscr{G}(\theta)(x)$ and $Z := (Y, X) \sim P_{\theta_0}$. Throughout, $P_N := N^{-1}\sum_{i=1}^N \delta_{Z_i}$ denotes the empirical measure induced by $Z_i \equiv (Y_i, X_i)_{i=1}^N$, which acts linearly on measurable functions $h : \mathbb{R} \times \mathcal{X} \to \mathbb{R}$ via

$$P_N(h) = \int_{\mathbb{R}\times\mathcal{X}} hdP_N = \frac{1}{N}\sum_{i=1}^N h(Z_i).$$

The following standard identities follow from the chain rule and will be used repeatedly in the proofs below – recall that ∇ and ∇^2 act on the θ-variable in E_D:

$$\begin{aligned}
-\ell(\theta, Z) &= \frac{1}{2}\big[Y - \mathscr{G}^X(\theta)\big]^2 = \frac{1}{2}\big[\mathscr{G}^X(\theta_0) + \varepsilon - \mathscr{G}^X(\theta)\big]^2,\\
-\nabla\ell(\theta, Z) &= \big[\mathscr{G}^X(\theta) - Y\big]\nabla\mathscr{G}^X(\theta),\\
-\nabla^2\ell(\theta, Z) &= \nabla\mathscr{G}^X(\theta)\nabla\mathscr{G}^X(\theta)^T + \big[\mathscr{G}^X(\theta) - Y\big]\nabla^2\mathscr{G}^X(\theta),\\
-E_{\theta_0}\big[\ell(\theta, Z)\big] &= \frac{1}{2} + \frac{1}{2}E^X[\mathscr{G}^X(\theta_0) - \mathscr{G}^X(\theta)]^2.
\end{aligned} \tag{3.39}$$

When no confusion can arise, we will suppress the second argument Z and write $\ell(\theta)$ for $\ell(\theta, Z)$. We also set $\bar{\ell}_N := \ell_N/N$.

By Condition 3.2.1, the matrix $-\nabla^2\bar{\ell}_N(\theta)$ is symmetric and by a standard inequality for eigenvalues due to Weyl and Condition 3.2.2, we have for any $\theta \in \mathcal{B}$ that

$$\begin{aligned}
\lambda_{min}\big[-\nabla^2\bar{\ell}_N(\theta)\big] &\geq \lambda_{min}\big(E_{\theta_0}\big[-\nabla^2\ell(\theta)\big]\big) - \big\|\nabla^2\bar{\ell}_N(\theta) - E_{\theta_0}\big[\nabla^2\ell(\theta)\big]\big\|_{op}\\
&\geq c_0D^{-\kappa_0} - \big\|\nabla^2\bar{\ell}_N(\theta) - E_{\theta_0}\big[\nabla^2\ell(\theta)\big]\big\|_{op}.
\end{aligned} \tag{3.40}$$

Hence we deduce

$$\begin{aligned}
&P_{\theta_0}^N\Big(\inf_{\theta\in\mathcal{B}} \lambda_{min}\big[-\nabla^2\ell_N(\theta, Z)\big] < Nc_0D^{-\kappa_0}/2\Big)\\
&\quad\leq P_{\theta_0}^N\Big(\big\|\nabla^2\bar{\ell}_N(\theta) - E_{\theta_0}\big[\nabla^2\ell(\theta)\big]\big\|_{op} \geq c_0D^{-\kappa_0}/2 \text{ for some } \theta \in \mathcal{B}\Big)\\
&\quad\leq P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\sup_{v:\|v\|_{\mathbb{R}^D}\leq 1}\Big|v^T\Big(\nabla^2\bar{\ell}_N(\theta) - E_{\theta_0}[\nabla^2\ell(\theta)]\Big)v\Big| \geq c_0D^{-\kappa_0}/2\Big).
\end{aligned} \tag{3.41}$$

The next step is to reduce the supremum over $\{v : \|v\|_{\mathbb{R}^D} \le 1\}$ to a suitable finite maximum over grid points v_i by a contraction argument (commonly used in high-dimensional probability, e.g., [132]). For $\rho > 0$, let $N(\rho)$ denote the minimal number of balls of $\|\cdot\|_{\mathbb{R}^D}$−radius ρ required to cover $\{v : \|v\|_{\mathbb{R}^D} \le 1\}$, and let $v_i, \|v_i\|_{\mathbb{R}^D} \le 1$, be the centre points of a minimal covering. Thus for any $v \in \mathbb{R}^D$ there exists an index i such that $\|v - v_i\|_{\mathbb{R}^D} \le \rho$. Hence, writing shorthand

$$M_\theta = \nabla^2 \bar{\ell}_N(\theta) - E_{\theta_0}[\nabla^2 \ell(\theta)], \quad \theta \in \mathcal{B},$$

we have by the Cauchy-Schwarz inequality and the symmetry of the matrix M_θ,

$$\begin{aligned} v^T M_\theta v &= v_i^T M_\theta v_i + (v - v_i)^T M_\theta v + v_i^T M_\theta (v - v_i) \\ &\le v_i^T M_\theta v_i + \|v - v_i\|_{\mathbb{R}^D} \|M_\theta v\|_{\mathbb{R}^D} + \|v - v_i\|_{\mathbb{R}^D} \|M_\theta v_i\|_{\mathbb{R}^D} \\ &\le v_i^T M_\theta v_i + 2\rho \sup_{v:\|v\|_{\mathbb{R}^D} \le 1} v^T M_\theta v. \end{aligned}$$

Choosing $\rho = \frac{1}{4}$ and taking suprema it follows that for any $\theta \in \mathcal{B}$,

$$\sup_{v:\|v\|_{\mathbb{R}^D} \le 1} v^T M_\theta v \le 2 \max_{i=1,\dots,N(1/4)} v_i^T M_\theta v_i. \tag{3.42}$$

Since the covering (v_i) is independent of θ, we can further estimate the right hand side of (3.41) by a union bound to the effect that

$$\begin{aligned} & P_{\theta_0}^N \Big(\sup_{\theta \in \mathcal{B}} \sup_{v:\|v\|_{\mathbb{R}^D} \le 1} \Big| v^T M_\theta v \Big| \ge c_0 D^{-\kappa_0}/2 \Big) \\ & \le N(1/4) \cdot \sup_{v:\|v\|_{\mathbb{R}^D} \le 1} P_{\theta_0}^N \Big(\sup_{\theta \in \mathcal{B}} \Big| v^T M_\theta v \Big| \ge c_0 D^{-\kappa_0}/4 \Big) \\ & \le N(1/4) \cdot \sup_{v:\|v\|_{\mathbb{R}^D} \le 1} \Big[P_{\theta_0}^N \Big(\sup_{\theta \in \mathcal{B}} \Big| P_N (g_{v,\theta} - g_{v,\theta_{0,D}}) \Big| \ge c_0 D^{-\kappa_0}/8 \Big) \\ & \qquad\qquad + P_{\theta_0}^N \big(\big| P_N (g_{v,\theta_{0,D}}) \big| \ge c_0 D^{-\kappa_0}/8 \big) \Big], \end{aligned} \tag{3.43}$$

where we have defined

$$g_{v,\theta}(\cdot) := v^T \Big(\nabla^2 \ell(\theta, \cdot) - E_{\theta_0}[\nabla^2 \ell(\theta)] \Big) v, \qquad v \in \mathbb{R}^D.$$

and where we recall that $\theta_{0,D}$ is the centre-point of the set $\mathcal{B}$ from (3.31). For the rest of the proof, we fix any $v \in \mathbb{R}^D$ with $\|v\|_{\mathbb{R}^D} \le 1$. Next, we use (3.39) to decompose the ‘uncentred’ part of $g_{v,\theta}$ as

$$\begin{aligned} & - v^T \nabla^2 \ell(\theta, Z) v \\ & = v^T \Big[\nabla \mathscr{G}^X(\theta) \nabla \mathscr{G}^X(\theta)^T + \big[\mathscr{G}^X(\theta) - \mathscr{G}^X(\theta_0) \big] \nabla^2 \mathscr{G}^X(\theta) \Big] v - \varepsilon v^T \nabla^2 \mathscr{G}^X(\theta) v \\ & =: \tilde{g}_{v,\theta}^I(X) + \varepsilon g_{v,\theta}^{II}(X), \end{aligned}$$

such that $g_{v,\theta}(z) = g^I_{v,\theta}(x) + \varepsilon g^{II}_{v,\theta}(x)$, where we have defined the centred version of $\tilde{g}^I_{v,\theta}$ as

$$g^I_{v,\theta}(x) = \tilde{g}^I_{v,\theta}(x) - E_{\theta_0}[\tilde{g}^I_{v,\theta}(X)], \quad x \in \mathcal{X}.$$

We can therefore bound the right hand side of (3.43) by

$$\begin{aligned}
N\Big(\frac{1}{4}\Big) \cdot \sup_{v:\|v\|_{\mathbb{R}^D}\le 1} \Bigg[& P^N_{\theta_0}\Big(\sup_{\theta\in\mathcal{B}} \Big|\frac{1}{N}\sum_{i=1}^N (g^I_{v,\theta} - g^I_{v,\theta_{0,D}})(X_i)\Big| \ge \frac{c_0 D^{-\kappa_0}}{16}\Big) \\
& + P^N_{\theta_0}\Big(\Big|\frac{1}{N}\sum_{i=1}^N g^I_{v,\theta_{0,D}}(X_i)\Big| \ge \frac{c_0 D^{-\kappa_0}}{16}\Big) \\
& + P^N_{\theta_0}\Big(\sup_{\theta\in\mathcal{B}} \Big|\frac{1}{N}\sum_{i=1}^N \varepsilon_i (g^{II}_{v,\theta} - g^{II}_{v,\theta_{0,D}})(X_i)\Big| \ge \frac{c_0 D^{-\kappa_0}}{16}\Big) \\
& + P^N_{\theta_0}\Big(\Big|\frac{1}{N}\sum_{i=1}^N \varepsilon_i g^{II}_{v,\theta_{0,D}}(X_i)\Big| \ge \frac{c_0 D^{-\kappa_0}}{16}\Big)\Bigg] =: N(1/4)\cdot(i + ii + iii + iv).
\end{aligned}$$

We now use empirical process techniques (Lemma 6.2.4 and also Hoeffding's inequality) to bound the preceding probabilities.

Terms i **and** ii**.** In order to apply Lemma 6.2.4 to term i, we require some preparations. By the definition of $\tilde{g}^I_{v,\theta}$ and of the operator norm $\|\cdot\|_{op}$, using the elementary identity $v^T(aa^T - bb^T)v = v^T(a+b)(a-b)^T v$ for any $v, a, b \in \mathbb{R}^D$ we have that for any $\theta, \bar\theta \in \mathcal{B}$,

$$\begin{aligned}
\|\tilde{g}^I_{v,\theta} - \tilde{g}^I_{v,\bar\theta}\|_\infty \le \Big\| & \big[\nabla\mathscr{G}(\theta)\nabla\mathscr{G}(\theta)^T + \big[\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\big]\nabla^2\mathscr{G}(\theta)\big] \\
& - \big[\nabla\mathscr{G}(\bar\theta)\nabla\mathscr{G}(\bar\theta)^T + \big[\mathscr{G}(\bar\theta) - \mathscr{G}(\theta_0)\big]\nabla^2\mathscr{G}(\bar\theta)\big]\Big\|_{L^\infty(\mathcal{X},\mathbb{R}^{D\times D})} \\
\le & \Big\|\big[\nabla\mathscr{G}(\theta) - \nabla\mathscr{G}(\bar\theta)\big]\big[\nabla\mathscr{G}(\theta) + \nabla\mathscr{G}(\bar\theta)\big]^T\Big\|_{L^\infty(\mathcal{X},\mathbb{R}^{D\times D})} \\
& + \Big\|\big[\mathscr{G}(\theta) - \mathscr{G}(\bar\theta)\big]\nabla^2\mathscr{G}(\theta)\Big\|_{L^\infty(\mathcal{X},\mathbb{R}^{D\times D})} \\
& + \Big\|\big[\mathscr{G}(\bar\theta) - \mathscr{G}(\theta_0)\big]\big[\nabla^2\mathscr{G}(\theta) - \nabla^2\mathscr{G}(\bar\theta)\big]\Big\|_{L^\infty(\mathcal{X},\mathbb{R}^{D\times D})} \\
\le & \, 2C_{\mathscr{G},D}|\theta - \bar\theta|_{E_D}
\end{aligned} \tag{3.44}$$

where Condition 3.2.1 furnishes the Lipschitz constant

$$C_{\mathscr{G},D} := c_2^2 D^{2\kappa_2}. \tag{3.45}$$

The hypothesis $\mathbf{r} \leq 1$ also implies that the $\tilde{g}^I_{v,\theta} - \tilde{g}^I_{v,\bar{\theta}}$ are uniformly bounded by $4C_{\mathscr{G},D}$. We then introduce the rescaled function class

$$h^I_\theta := \frac{g^I_{v,\theta} - g^I_{v,\theta_{0,D}}}{16C_{\mathscr{G},D}}, \qquad \mathcal{H}^I = \{h^I_\theta : \theta \in \mathcal{B}\},$$

which has envelope and variance proxy bounded as

$$\sup_{\theta\in\mathcal{B}} \|h^I_\theta\|_\infty \leq 1/4 \equiv U, \qquad \sup_{\theta\in\mathcal{B}} \big(E_{\theta_0}\big[h^I_\theta(X)^2\big]\big)^{\frac{1}{2}} \leq 1/4 \equiv \sigma. \tag{3.46}$$

Next, if

$$d_2^2(\theta,\bar{\theta}) = E_{\theta_0}\big[(h^I_\theta(X) - h^I_{\bar{\theta}}(X))^2\big], \;\; d_\infty(\theta,\bar{\theta}) = \|h^I_\theta - h^I_{\bar{\theta}}\|_\infty, \qquad \theta,\bar{\theta} \in \mathcal{B},$$

then using (3.44) we have that

$$d_2(\theta,\bar{\theta}) \leq d_\infty(\theta,\bar{\theta}) \leq |\theta - \bar{\theta}|_{E_D}, \qquad \theta,\bar{\theta} \in \mathcal{B}.$$

Thus for any $\rho \in (0,1)$, using Proposition 4.3.34 in [61], we obtain that

$$N\big(\mathcal{H}^I, d_2, \rho\big) \leq N\big(\mathcal{H}^I, d_\infty, \rho\big) \leq N\big(\mathcal{B}, |\cdot|_{E_D}, \rho\big) \leq (3/\rho)^D. \tag{3.47}$$

For any $A \geq 2$ we have

$$\int_0^1 \log(A/x)dx = \log(A) + 1, \qquad \int_0^1 \sqrt{\log(A/x)}dx \leq \frac{2\log A}{2\log A - 1}\sqrt{\log(A)},$$

[see p.190 of [61] for the latter standard inequality], and hence, using $A = 3$, we can respectively bound the L^∞ and L^2 metric entropy integrals of $\mathcal{H}^I$ by

$$\begin{aligned} \mathcal{J}_\infty(\mathcal{H}^I) &= \int_0^{4U} \log N(\mathcal{H}^I, d_\infty, \rho)d\rho \lesssim D, \\ J_2(\mathcal{H}^I) &\leq \int_0^{4\sigma} \sqrt{\log N(\mathcal{H}^I, d_2, \rho)}d\rho \lesssim \sqrt{D}. \end{aligned}$$

Now, an application of Lemma 6.2.4 below implies that for any $x \geq 1$ and some universal constant $L' > 0$, we have that

$$P^N_{\theta_0}\Big(\sup_{\theta\in\mathcal{B}} \frac{1}{\sqrt{N}}\Big|\sum_{i=1}^N h^I_\theta(X_i)\Big| \geq L'\Big[\sqrt{D} + \sqrt{x} + (D + x)/\sqrt{N}\Big]\Big) \leq 2e^{-x}. \tag{3.48}$$

We also have by the definition of $g^I_{v,\theta_{0,D}}$ that

$$\|g^I_{v,\theta_{0,D}}\|_\infty \leq 2\|\tilde{g}^I_{v,\theta_{0,D}}\|_\infty \leq 2C_{\mathscr{G},D},$$

with $C_{\mathscr{G},D}$ from (3.45), and hence by Hoeffding's inequality (Theorem 3.1.2 in [61])

$$ii \le 2\exp\Big(-\frac{N(c_0D^{-\kappa_0})^2}{512C_{\mathscr{G},D}^2}\Big). \tag{3.49}$$

Thus from (3.45), $D \le \mathcal{R}_N$ and choosing $x = 4\mathcal{R}_N$ we have

$$L'\Big[\sqrt{D}+\sqrt{x}+(D+x)/\sqrt{N}\Big] \le \frac{c_0D^{-\kappa_0}\sqrt{N}}{256C_{\mathscr{G},D}}, \quad 4\mathcal{R}_N \le \frac{N(c_0D^{-\kappa_0})^2}{512C_{\mathscr{G},D}^2},$$

whenever $C>0$ in (3.36) is small enough. Therefore, combining (3.48) and (3.49), and using the definitions of the term i and of h_θ^I, we obtain

$$ii + i \le 2e^{-4\mathcal{R}_N} + P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\frac{1}{\sqrt{N}}\Big|\sum_{i=1}^N h_\theta^I(X_i)\Big| \ge \frac{c_0D^{-\kappa_0}\sqrt{N}}{256C_{\mathscr{G},D}}\Big) \le 4e^{-4\mathcal{R}_N}. \tag{3.50}$$

Terms iii **and** iv**.** Let us now treat the empirical process indexed by the functions $\{g_{v,\theta}^{II} : \theta \in \mathcal{B}\}$. Since $\|v\|_{\mathbb{R}^D} \le 1$, we have for $c_{\mathscr{G},D} = \sqrt{C_{\mathscr{G},D}} = c_2D^{\kappa_2}$, and any $\theta, \bar\theta \in \mathcal{B}$,

$$\|g_{v,\theta}^{II} - g_{v,\bar\theta}^{II}\|_\infty \le \|\nabla^2\mathscr{G}(\theta) - \nabla^2\mathscr{G}(\bar\theta)\|_{L^\infty(\mathcal{X},\mathbb{R}^{D\times D})} \le c_{\mathscr{G},D}|\theta-\bar\theta|_{E_D} \le c_{\mathscr{G},D},$$

using also $\mathbf{r} \le 1$ to obtain the envelope bound in the last inequality. Now the rescaled function class

$$h_\theta^{II} := \frac{g_{v,\theta}^{II} - g_{v,\theta_{0,D}}^{II}}{4c_{\mathscr{G},D}}, \quad \mathcal{H}^{II} = \{h_\theta^{II} : \theta \in \mathcal{B}\},$$

admits envelopes

$$\sup_{\theta\in\mathcal{B}}\|h_{v,\theta}^{II}\|_\infty \le 1/4 \equiv U, \quad \sup_{\theta\in\mathcal{B}}\big(E_{\theta_0}\big[h_{v,\theta}^{II}(X)^2\big]\big)^{\frac12} \le 1/4 \equiv \sigma,$$

and defining

$$d_2^2(\theta,\bar\theta) := E_{\theta_0}\big[(h_{v,\theta}^{II}(X) - h_{v,\bar\theta}^{II}(X))^2\big], \quad d_\infty(\theta,\bar\theta) = \|h_{v,\theta}^{II} - h_{v,\bar\theta}^{II}\|_\infty, \quad \theta,\bar\theta\in\mathcal{B}$$

we have

$$d_2(\theta,\bar\theta) \le d_\infty(\theta,\bar\theta) \le |\theta-\bar\theta|_{E_D}, \quad \theta,\bar\theta \in \mathcal{B}.$$

Therefore, just as with the entropy integral bounds obtained for term i above, we have $N\big(\mathcal{H}^{II}, d_2, \rho\big) \le (3/\rho)^D$ and thus, by Lemma 6.2.4 below,

$$P_{\theta_0}^N\Big(\sup_{\theta\in\mathcal{B}}\frac{1}{\sqrt{N}}\Big|\sum_{i=1}^N \varepsilon_i h_\theta^{II}(X_i)\Big| \ge L'\Big[\sqrt{D}+\sqrt{x}+(D+x)/\sqrt{N}\Big]\Big) \le 2e^{-x}, \quad x \ge 1. \tag{3.51}$$

Moreover, by the hypotheses, $\|g^{II}_{v,\theta}\|_\infty \le c_{\mathscr{G},D}$, and hence, invoking the Bernstein inequality (6.48) with $U = \sigma \equiv c_{\mathscr{G},D}$, we obtain that

$$P^N_{\theta_0}\Big(\Big|\frac{1}{\sqrt{N}}\sum_{i=1}^N \varepsilon_i g^{II}_{v,\theta}(X_i)\Big| \ge c_{\mathscr{G},D}\big(\sqrt{2x} + \frac{x}{3\sqrt{N}}\big)\Big) \le 2e^{-x}, \quad x > 0. \tag{3.52}$$

Choosing $x = 4\mathcal{R}_N$ in the preceding displays, we obtain that for $C > 0$ small enough in (3.36), any $D \le \mathcal{R}_N$, and by similar calculations as before (3.50),

$$\begin{aligned} iii + iv \le P^N_{\theta_0}\Big(\sup_{\theta\in\mathcal{B}} \frac{1}{\sqrt{N}}\Big|\sum_{i=1}^N \varepsilon_i h^{II}_\theta(X_i)\Big| \ge \frac{c_0 D^{-\kappa_0}\sqrt{N}}{96 m_2}\Big) \\ + P^N_{\theta_0}\Big(\Big|\frac{1}{\sqrt{N}}\sum_{i=1}^N \varepsilon_i g^{II}_{v,\theta}(X_i)\Big| \ge \frac{c_0 D^{-\kappa_0}\sqrt{N}}{16}\Big) \le 4e^{-4\mathcal{R}_N}. \end{aligned} \tag{3.53}$$

Combining the terms. By combining the bounds (3.41), (3.43), (3.50), (3.53) and using that $N(1/4) \le 9^D \le e^{3D}$ (cf. Proposition 4.3.34 in [61]) we obtain for $D \le \mathcal{R}_N$ and our choice of $\mathcal{R}_N$ with C small enough the final bound

$$\begin{aligned} P^N_{\theta_0}\Big(\inf_{\theta\in\mathcal{B}} \lambda_{min}\big(-\nabla^2 \ell_N(\theta, Z)\big) < N c_0 D^{-\kappa_0}/2\Big) &\le N(1/4)\cdot(i + ii + iii + iv) \\ &\le 8e^{-\mathcal{R}_N}, \end{aligned}$$

completing the proof of (3.37). The bound (3.38) is proved in a similar (in fact simpler) way and left to Ex. 3.4.3. □

3.3 Information operators for elliptic PDEs

We now return to the model examples with forward maps (2.5), (2.8) arising from the PDE (1.2) with elliptic operator $\mathscr{L}_{f_\theta}$ from (1.4), (1.3), respectively. For both PDEs we will derive the linearisation $\mathbb{I}_\theta$ from Condition 3.1.1, its adjoint for natural infinite-dimensional tangent spaces H, and the resulting information operator from Definition 3.1.2. Recall that for these examples we take $V = W = \mathbb{R}$ and $\zeta = \lambda$ equal to Lebesgue measure on the bounded smooth domain $\mathcal{X} = \mathcal{Z}$ in the general setting of the preceding subsection.

The main idea behind checking Condition 3.1.1 is to write

$$\mathscr{G}(\theta_0 + h) - \mathscr{G}(\theta_0) - D\mathscr{G}_{\theta_0}[h], \quad h \in H,$$

as the solution of particular *inhomogeneous* elliptic PDE for appropriate choice of $D\mathscr{G}_{\theta_0}$, and to use regularity estimates for the solution operators $\mathscr{L}^{-1}_{f_\theta}$ to verify (3.1).

3.3.1 Schrödinger equation

For the forward map $\mathscr{G}(\theta) = u_\theta$ from (2.5) arising with the Schrödinger operator $\mathscr{L}_{f_\theta}$ from (1.4), let us verify Condition 3.1.1 with $\Theta = H = H^\xi(\mathcal{X}), \xi > d/2$, and

$$\mathbb{I}_\theta[h] \equiv D\mathscr{G}_\theta[h] = \mathscr{L}_{f_\theta}^{-1}[u_\theta e^\theta h], \quad h \in H. \tag{3.54}$$

where $\mathscr{L}_{f_\theta}^{-1}$ is the inverse Schrödinger operator for Dirichlet boundary conditions (given in (6.10) for $\gamma = 1/2, V = f_\theta$). We see from (6.12), (6.30), (6.4) that $h \mapsto \mathbb{I}_\theta(h)$ is a bounded and continuous operator on L^2_λ for every fixed θ. To check that $\mathbb{I}_\theta$ is the linearisation $\mathscr{G}$ at a point $\theta \in \Theta$, consider a perturbation $\theta + h$ in any direction h in the tangent space H. Then on $\mathcal{X}$,

$$\begin{aligned}
\mathscr{L}_{f_{\theta+h}}[u_{\theta+h} - u_\theta - \mathbb{I}_\theta[h]] &= 0 + (\mathscr{L}_{f_\theta} - \mathscr{L}_{f_{\theta+h}})[u_\theta + \mathbb{I}_\theta[h]] - \mathscr{L}_{f_\theta}[\mathbb{I}_\theta[h]] \\
&= (f_{\theta+h} - f_\theta)u_\theta + (f_{\theta+h} - f_\theta)\mathbb{I}_\theta[h] - u_\theta e^\theta h \\
&= (e^{\theta+h} - e^\theta - e^\theta h)u_\theta - (e^{\theta+h} - e^\theta)\mathbb{I}_\theta[h] \equiv R(h)
\end{aligned}$$

while at the boundary $\partial\mathcal{X}$ we necessarily have $u_{\theta+h} - u_\theta - \mathbb{I}_\theta[h] = 0$. Hence for every $h \in H$,

$$\mathscr{G}(\theta + h) - \mathscr{G}(\theta) - D\mathscr{G}_\theta[h] = u_{\theta+h} - u_\theta - \mathbb{I}_\theta[h] = \mathscr{L}_{f_{\theta+h}}^{-1}[R(h)]. \tag{3.55}$$

Now using the PDE estimates (6.12), (6.30) both with constant c depending only on $\mathcal{X}$, and standard properties of the exponential map for bounded θ, $\|\theta\|_\infty \lesssim \|\theta\|_{H^\xi}$ (in view of (6.4)), we see

$$\begin{aligned}
\|\mathscr{G}(\theta + h) - \mathscr{G}(\theta) - D\mathscr{G}_\theta[h]\|_{L^2} &= \|\mathscr{L}_{f_{\theta+h}}^{-1}[R(h)]\|_{L^2} \lesssim \|R(h)\|_{L^2} \\
&\lesssim \|(e^{\theta+h} - e^\theta - e^\theta h)u_\theta\|_{L^2} + \|(e^{\theta+h} - e^\theta)\mathbb{I}_\theta[h]\|_{L^2} \\
&\lesssim \|h\|_\infty^2 + \|h\|_\infty\|\mathbb{I}_\theta[h]\|_{L^2} \lesssim \|h\|_\infty^2.
\end{aligned}$$

Summarising what precedes we have proved:

Theorem 3.3.1. *Let $\Theta = H^\xi(\mathcal{X})$ for some $\xi > d/2$. The forward map $\mathscr{G}$ from (2.5) satisfies Condition 3.1.1 for any $\theta = \theta_0 \in \Theta$, tangent space $H = H^\xi(\mathcal{X})$, operator $\mathbb{I}_\theta$ from (3.54) and $\rho_\theta(h) \leq C\|h\|_\infty^2$ with some constant C that depends only on θ and $\mathcal{X}$.*

In fact for the forward map $\mathscr{G}$ from (2.5) can be show to verify (3.1) with weaker norms than $\|\cdot\|_\infty$ measuring the size of the perturbation h (see, e.g., [94]), but for the present purposes the above proposition is sufficient.

If we choose as tangent space $H = H^\xi(\mathcal{X})$ or just $H = H^\xi_c(\mathcal{X})$, then its completion in $L^2(\mathcal{X})$ equals $\bar{H} = L^2(\mathcal{X})$ itself, and the adjoint operator of $\mathbb{I}_\theta$ then takes a particularly simple form

$$\mathbb{I}_\theta^* g = u_\theta e^\theta \mathscr{L}_{f_\theta}^{-1}[g]. \tag{3.56}$$

Indeed, using that $\mathscr{L}_{f_\theta}^{-1}$ is self-adjoint for the $L^2_\lambda(\mathcal{X})$-inner product (cf. (6.10)),

$$\langle \mathbb{I}_\theta[h], g\rangle_{L^2} = \langle \mathscr{L}_{f_\theta}^{-1}[u_\theta e^\theta h], g\rangle_{L^2} = \langle h, u_\theta e^\theta \mathscr{L}_{f_\theta}^{-1}[g]\rangle_{L^2} = \langle h, \mathbb{I}_\theta^* g\rangle_{L^2}.$$

One shows further that the information operator $\mathbb{I}_\theta^*\mathbb{I}_\theta$ arising from (2.5) as in Definition 3.1.2 is a self-adjoint and *compact* operator on L^2 (see Ex. 3.4.4).

3.3.2 Diffusion equation

In the following theorem $\mathscr{L}_{f_\theta}^{-1}$ denotes the inverse of the divergence form operator $\mathscr{L}_{f_\theta}$ from (1.3) for Dirichlet boundary conditions – see (6.10) (with $\gamma = f_\theta$ and $V = 0$).

Theorem 3.3.2. *Let $\Theta = H^\beta(\mathcal{X}), \beta > 1 + d/2$. The forward map $\mathscr{G}(\theta) = u_\theta = u_{f_\theta}$ from (2.8) satisfies Condition 3.1.1 for any $\theta = \theta_0 \in \Theta$, tangent space $H = H^\beta(\mathcal{X})$, operator*

$$\mathbb{I}_\theta(h) \equiv D\mathscr{G}_\theta[h] = -\mathscr{L}_{f_\theta}^{-1}[\nabla\cdot(e^\theta h \nabla u_{f_\theta})], \quad h \in H, \tag{3.57}$$

and $\rho_\theta(h) \le C\|h\|_\infty^2$ for some constant $C = C(\mathcal{X}, f_{min}, \theta)$.

Proof. Fix $\theta \in \Theta$, let $h \in H$ be arbitrary, and write u_θ for u_{f_θ} in this proof. We first observe that

$$\begin{aligned}\mathscr{L}_{f_{\theta+h}}(u_{\theta+h} - u_\theta - \mathbb{I}_\theta(h)) &= g - (\mathscr{L}_{f_{\theta+h}} - \mathscr{L}_{f_\theta})(u_\theta + \mathbb{I}_\theta(h)) - \mathscr{L}_{f_\theta}(u_\theta + \mathbb{I}_\theta(h)) \\ &= -\big(\nabla\cdot((e^{\theta+h} - e^\theta)\nabla(u_\theta + \mathbb{I}_\theta(h))\big) + \nabla\cdot(e^\theta h \nabla u_\theta) \\ &= -\nabla\cdot((e^{\theta+h} - e^\theta - e^\theta h)\nabla u_\theta) - \nabla\cdot((e^{\theta+h} - e^\theta)\nabla \mathbb{I}_\theta(h)) \\ &\equiv R_1(h) + R_2(h),\end{aligned}$$

as well as $u_{\theta+h} - u_\theta - \mathbb{I}_\theta(h) = 0$ on $\partial\mathcal{X}$. Therefore we can write

$$\begin{aligned}\|\mathscr{G}(\theta+h) - \mathscr{G}(\theta) - D\mathscr{G}_\theta[h]\|_{L^2} &= \|u_{\theta+h} - u_\theta - \mathbb{I}_\theta[h]\|_{L^2} \\ &= \|\mathscr{L}_{f_{\theta+h}}^{-1}\big[R_1(h) + R_2(h)\big]\|_{L^2} \\ &\le \|\mathscr{L}_{f_{\theta+h}}^{-1}[R_1(h)]\|_{L^2} + \|\mathscr{L}_{f_{\theta+h}}^{-1}[R_2(h)]\|_{L^2}.\end{aligned}$$

We now show that the r.h.s. is $O(\|h\|_\infty^2)$. For the first term we have from the regularity estimate (6.36) with $\gamma = f_\theta, V = 0$ and Lipschitz constant $\bar{c} = \bar{c}(\mathcal{X}, f_{min})$, as well as the divergence theorem (6.7),

$$\begin{aligned}\|\mathscr{L}_{f_{\theta+h}}^{-1}\big[R_1(h)\big]\|_{L^2} &\le \bar{c}\|\nabla\cdot((e^{\theta+h} - e^\theta - e^\theta h)\nabla u_\theta)\|_{(H_0^1)^*} \\ &\lesssim \|(e^{\theta+h} - e^\theta - e^\theta h)\nabla u_\theta)\|_{L^2} \lesssim \|h\|_\infty^2\end{aligned}$$

using also that $u_\theta \in H^1$ under the maintained hypotheses (e.g., Proposition 6.1.5). For the second term we have similarly, using (6.36) and (6.8) this time,

$$\begin{aligned}
\|\mathscr{L}_{f_{\theta+h}}^{-1}[R_2(h)]\|_{L^2} &\lesssim \left\|\nabla \cdot ((e^{\theta+h} - e^\theta)\nabla \mathbb{I}_\theta(h))\right\|_{(H_0^1)^*} \\
&\lesssim \|h\|_\infty \|\nabla \mathscr{L}_{f_\theta}^{-1}[\nabla \cdot (e^\theta h \nabla u_\theta))]\|_{L^2} \\
&\lesssim \|h\|_\infty \|\nabla \cdot (e^\theta h \nabla u_\theta)\|_{(H_0^1)^*} \\
&\lesssim \|h\|_\infty^2 \|u_\theta\|_{H^1} \lesssim \|h\|_\infty^2,
\end{aligned}$$

The last two inequalities also imply the continuity of $\mathbb{I}_\theta$ as an operator on L^2, completing the proof. $\square$

Just as with the Schrödinger equation, if we take the tangent space $H = H^\beta$, or just H_c^β, then its closure $\bar{H}$ for the L^2-norm equals all of L^2 and the adjoint in (3.3) is computed for the standard L^2-inner product.

Proposition 3.3.3. *In the setting of Theorem 3.3.2, the adjoint $\mathbb{I}_\theta^* : L^2_\lambda(\mathcal{X}) \to L^2_\lambda(\mathcal{X})$ of $\mathbb{I}_\theta : L^2_\lambda(\mathcal{X}) \to L^2_\lambda(\mathcal{X})$ is given by*

$$\mathbb{I}_\theta^*[g] = e^\theta \nabla u_\theta \cdot \nabla \mathscr{L}_{f_\theta}^{-1}[g], \quad g \in L^2_\lambda(\mathcal{X}). \tag{3.58}$$

Proof. Since $\mathbb{I}_\theta$ from (3.57) defines a bounded linear operator on the Hilbert space $L^2_\lambda(\mathcal{X})$, a unique adjoint operator I_θ^* exists by the Riesz-representation theorem. Let us first show that

$$\langle h, (I_\theta^* - \mathbb{I}_\theta^*)g\rangle_{L^2} = 0, \quad \forall h, g \in C_0^\infty(\mathcal{X}). \tag{3.59}$$

Indeed, since $\mathscr{L}_{f_\theta}^{-1}$ from (6.10) is self-adjoint on L^2_λ and maps into $H_0^1(\mathcal{X})$, we can apply Theorem 3.3.2 and the divergence theorem (6.7) with vector field $e^\theta h \nabla u_\theta$ to deduce

$$\begin{aligned}
\langle h, I_\theta^* g\rangle_{L^2(\mathcal{X})} = \langle \mathbb{I}_\theta h, g\rangle_{L^2(\mathcal{X})} &= -\langle \mathscr{L}_{f_\theta}^{-1}[\nabla \cdot (e^\theta h \nabla u_\theta)], g\rangle_{L^2(\mathcal{X})} \\
&= -\int_{\mathcal{X}} [\nabla \cdot (e^\theta h \nabla u_\theta)] \mathscr{L}_{f_\theta}^{-1}[g] d\lambda \\
&= \int_{\mathcal{X}} h e^\theta \nabla u_\theta \cdot \nabla \mathscr{L}_{f_\theta}^{-1}[g] d\lambda = \langle h, \mathbb{I}_\theta^* g\rangle_{L^2(\mathcal{X})},
\end{aligned}$$

so that (3.59) follows. Since $C_0^\infty(\mathcal{X})$ is dense in $L^2_\lambda(\mathcal{X})$ and since $I_\theta^*, \mathbb{I}_\theta^*$ are continuous on $L^2_\lambda(\mathcal{X})$ (by construction in the former case and by (6.13) and Proposition 6.1.5 in the latter case), the identity (3.59) extends to all $g \in L^2_\lambda(\mathcal{X})$ and hence $I_\theta^* = \mathbb{I}_\theta^*$, as desired. $\square$

From the previous two results one can define the information operator $\mathbb{I}_\theta^* \mathbb{I}_\theta$ acting continuously (in fact compactly, see Ex. 3.4.4) on L^2.

3.3.3 Injectivity and local identifiability

As mentioned at the outset of this chapter, the operators $\mathbb{I}_{\theta_0}, \mathbb{I}^*_{\theta_0}, \mathbb{I}^*_{\theta_0}\mathbb{I}_{\theta_0}$ encode local identifiability properties of a statistical model $\{dP_\theta : \theta \in \Theta\}$ at $\theta_0 \in \Theta$. In particular Theorem 3.1.5 implies that $\sqrt{N}$-consistent inference on functionals $\langle \theta, \psi \rangle_{L^2}$ at θ is a fortiori only possible if $\mathbb{I}_\theta$ is injective along tangent space directions ψ: If ψ is to lie in the range $R(\mathbb{I}^*_\theta)$ of the adjoint of $\mathbb{I}^*_\theta$, then it cannot lie in the kernel of $\mathbb{I}_\theta$ in view of (3.15). So it is natural to investigate the injectivity properties of $\mathbb{I}_\theta$ first, and in doing that we can also shed some light on the basic requirement (3.30) necessary to verify the local curvature hypothesis underpinning Theorem 3.2.3.

Schrödinger equation

Let us start again with the example of the Schrödinger equation. We show that the operator $\mathbb{I}_\theta$ given in Theorem 3.3.1 is injective on all of $L^2(\mathcal{X})$: We have from (3.54) and Proposition 6.1.4 (with $\gamma = 1/2, V = f_\theta$) that

$$\|\mathbb{I}_\theta h\|_{L^2} = \|\mathscr{L}_{f_\theta}^{-1}[u_\theta e^\theta h]\|_{L^2} \gtrsim \|u_\theta e^\theta h\|_{(H^2_0)^*} \tag{3.60}$$

for any $h \in L^2_\lambda(\mathcal{X})$. Under the hypotheses of Theorem 3.3.1, the function $u_\theta e^\theta$ is bounded away from zero throughout $\mathcal{X}$ and contained in $H^2(\mathcal{X})$ (cf. (6.29) and Proposition 6.1.5). Thus for any $\bar\varphi \in C_c^\infty(\mathcal{X})$ the function $\varphi = \bar\varphi/(e^\theta u_\theta)$ belongs to $H^2_0(\mathcal{X})$ and hence (3.60) implies that $\int \bar\varphi h = 0$ for any such $\bar\varphi$ whenever $\mathbb{I}_\theta h = 0$. But this implies $h = 0$ almost everywhere on $\mathcal{X}$ and hence that $\mathbb{I}_\theta$ is injective on $L^2(\mathcal{X})$.

What precedes implies as well that the information operator arising from $\mathbb{I}_\theta$ is injective on $L^2(\mathcal{X})$. Indeed if $\mathbb{I}^*_\theta \mathbb{I}_\theta h = 0$ then

$$0 = \langle h, \mathbb{I}^*_\theta \mathbb{I}_\theta h \rangle_{L^2} = \|\mathbb{I}_\theta h\|^2_{L^2} \tag{3.61}$$

and thus $\mathbb{I}_\theta h = 0$. We further notice that (3.60) verifies (3.30) for $\|\cdot\|_+ = \|\cdot\|_{(H^2_0)^*}$ which readily gives a recipe to check the local curvature Condition (3.28) if we can adapt the basis E_D to the dual norm $(H^2_0)^*$. This will be pursued further in Chapter 5.

Diffusion equation

For the divergence form equation, the injectivity of $\mathbb{I}_\theta$ on $L^2(\mathcal{X})$ is a more subtle question. For the operator $\mathbb{I}_\theta$ from Theorem 3.3.2 we can use Proposition 6.1.4 (with $\gamma = f_\theta, V = 0$) to obtain

$$\|\mathbb{I}_\theta h\|_{L^2} \gtrsim \|\nabla \cdot (e^\theta h \nabla u_{f_\theta})\|_{(H^2_0)^*}, \ h \in L^2. \tag{3.62}$$

Hence if $\mathbb{I}_\theta h = 0$ then by testing against all smooth $\varphi \in C_c^\infty \subset H_0^2$ we can deduce that also $\nabla \cdot (e^\theta h \nabla u_{f_\theta}) = 0$. Whether this implies that h vanishes depends on the precise form of the PDE (1.2).

If we impose the natural hypothesis (2.16) from earlier ensuring injectivity of the non-linear map $\mathscr{G}$, then Lemma 2.1.6 does imply that $\mathbb{I}_\theta$ (and as in (3.61) also $\mathbb{I}_\theta^* \mathbb{I}_\theta$) are injective on the space $H_0^1(\mathcal{X})$. A slight modification of this argument combined with Lemma 2.1.6 will allow us to verify the curvature Condition (3.28) for appropriate choice of $E_D \subset H_0^1$ as we will see in Theorem 5.3.2 below.

But on the other hand, Ex. 3.4.5 gives an example where $\mathbb{I}_\theta$ does satisfy (2.16) but is *not* injective on *all* of $L^2(\mathcal{X})$. This leads to information-theoretic obstructions in view of Theorem 3.1.5 and since the kernel in (3.15) is calculated on $L^2(\mathcal{X})$ and not on $H_0^1(\mathcal{X})$ – see Chapter 4.2 for more on this.

In some situations, however, $\mathbb{I}_\theta$ can be injective on all of $L^2(\mathcal{X})$, as in the following model example for the standard Laplacian $\theta = 0$ (and $f_{min} = 0$ in slight abuse of notation), which will also be of interest again in Chapter 4.2.

Proposition 3.3.4. *Let $\mathcal{X}$ equal to the unit disk in $\mathbb{R}^2$ centred at $(0,0)$ and let $\mathbb{I}_\theta$ be as in Theorem 3.3.2 where $\mathscr{G}$ from (2.8) arises with source $g = 2$ and boundary values $h = (|\cdot|_{\mathbb{R}^2}^2 - 1)/2$ $(= 0$ on $\partial\mathcal{X})$. Then for $f_{min} = 0, \theta = 0$, the map $\mathbb{I}_0 : L^2(\mathcal{X}) \to L^2(\mathcal{X})$ is injective.*

Proof. Suppose $\mathbb{I}_0(f) = 0$ for $f \in L^2(\mathcal{X})$. Then for any $g \in C^\infty(\mathcal{X})$ we have by Proposition 3.3.3

$$0 = \langle \mathbb{I}_0 f, g \rangle_{L^2(\mathcal{X})} = \langle f, \mathbb{I}_0^* g \rangle_{L^2(\mathcal{X})} = \langle f, X\mathscr{L}_1^{-1}[g] \rangle_{L^2(\mathcal{X})} \tag{3.63}$$

with vector field

$$X = \nabla u_1 \cdot \nabla(\cdot) = x_1 \partial x_1 + x_2 \partial x_2, \ (x_1, x_2) \in \mathcal{X}.$$

Choosing $g = \Delta \bar{g}$ for any smooth $\bar{g}$ of compact support we deduce that

$$\int_{\mathcal{X}} X(\bar{g}) f d\lambda = 0, \quad \forall \bar{g} \in C_c^\infty(\mathcal{X}), \tag{3.64}$$

and we now show that this implies $f = 0$. A somewhat informal dynamical argument would say that (3.64) asserts that $f d\lambda$ is an invariant density under the flow of X. Since the flow of X in backward time has a sink at the origin, the density can only be supported at $(x_1, x_2) = 0$ and thus $f = 0$.

One can give a distributional argument as follows. Suppose we consider polar coordinates $(r, \vartheta) \in (0,1) \times S^1$ and functions $\bar{g}$ of the form $\phi(r)\psi(\vartheta)$, where $\phi \in C_0^\infty(0,1)$ and $\psi \in C^\infty(S^1)$. In polar coordinates $X = r\partial_r$ and hence we may write (3.64) as

$$\int_0^1 \left(r^2 \left(\int_0^{2\pi} f(r,\vartheta)\psi(\vartheta)\, d\vartheta \right) \partial_r \phi \right) dr = 0. \tag{3.65}$$

By Fubini's theorem, for each ψ we have an integrable function

$$F_\psi(r) := \int_0^{2\pi} f(r,\vartheta)\psi(\vartheta)\,d\vartheta$$

and thus $r^2 F_\psi$ defines an integrable function on $(0,1)$ whose distributional derivative satisfies $\partial_r(r^2 F_\psi) = 0$ by virtue of (3.65). Thus $r^2 F_\psi = c_\psi$ (using that a distribution on $(0,1)$ with zero derivative must be a constant). Now consider $\psi \in C^\infty(S^1)$ also as a function in $L^2(\mathcal{X})$ and compute the pairing

$$(f,\psi)_{L^2(\mathcal{X})} = \int_0^1 rF_\psi(r)\,dr = c_\psi \int_0^1 r^{-1}\,dr = \pm\infty$$

unless $c_\psi = 0$. Thus $f = 0$. □

By perturbation and the Morse lemma, one can show further that the preceding injectivity result extends at least to $\mathbb{I}_\theta$ for an appropriate neighbourhood of θ's near the standard Laplacian $\theta = 0$, see Ex. 3.4.6.

3.4 Notes

3.4.1 Exercises

Exercise 3.4.1. Show that Theorem 3.1.4 remains valid if $\mathbb{I}_\theta$ is not injective by repeating its proof on the quotient space of $\tilde{H}$ for the kernel of $\mathbb{I}_\theta$.

Exercise 3.4.2. Show that Theorem 3.1.4 and (3.19) remain valid in the augmented model from the proof of Theorem 3.1.5. [Hint: Proceed as in the proof of Theorem 3.1.4 but apply Theorem 3.11.5 from [130] with the LAN expansion of the augmented model.]

Exercise 3.4.3. Provide a proof of (3.38). [Follow the proof of (3.37) in Lemma 3.2.3, or see Lemma 3.4 in [102].]

Exercise 3.4.4. Show that the information operators $\mathbb{I}_\theta^*\mathbb{I}_\theta$ from Definition 3.1.2 arising from $\mathbb{I}_\theta$ given in Theorems 3.3.1 and 3.3.2, respectively, are *compact* operators on $L^2_\lambda(\mathcal{X})$.

Exercise 3.4.5. Consider $\mathscr{G}, h, g$ as in Proposition 3.3.4 but now with $\mathcal{X}$ a smooth bounded domain in $\mathbb{R}^2$ that is *separated away from the origin.* Show that ∇u_1 does not vanish on $\mathcal{X}$ and hence satisfies (2.16), but that $\mathbb{I}_0$ is *not* injective on $L^2_\lambda(\mathcal{X})$. [Hint: See Sec. 3.4.1 in [96].]

Exercise 3.4.6. Show that Proposition 3.3.4 extends to all θ such that $\|\theta\|_{H^\beta} < \eta$ for $\beta > 1 + d/2$ and $\eta > 0$ small enough. [Hint: Show first by perturbation that $\|u_{f_\theta} - u_1\|_{C^1} \to 0$ as $\eta \to 0$ and that hence ∇u_θ can only vanish at a single point for η sufficiently small. Then repeat the proof of Proposition 3.3.4.]

3.4.2 Remarks and comments

The results from Section 3.1 constitute by now classical material from semi-parametric statistics, we refer to [126, 127] where these ideas are laid out in general likelihood models of which (1.11) is just a special case. The preceding references develop a formal 'score operator calculus' in general statistical models that are 'differentiable in quadratic mean (DQM)', which for our (1.11) leads initially to slightly different expressions for $\mathbb{I}_\theta$ involving a further 'projection' step onto the regression residuals $Y - \mathscr{G}(\theta)(X) = \varepsilon$. This is easily reconciled with the form of $\mathbb{I}_\theta$ given here as discussed in [96]. Since the LAN property in Theorem 3.1.3 with LAN norm $\|\mathbb{I}_{\theta_0}(\cdot)\|_{L^2}$ can be proved directly in our Gaussian regression model, we can avoid introducing the DQM property altogether. This expedites the proofs of Theorems 3.1.4 and 3.1.5, but we again refer to [96] for a discussion of how the model (1.11) does fit into the abstract framework from [126, 127].

The connection between local average curvature of the log-likelihood function and the information matrix is well-known in 'parametric' finite-dimensional statistical models – the idea to exploit stability estimates (3.30) in conjunction with high-dimensional concentration of measure arguments in Theorem 3.2.3 has been developed in this form in [102], see also [23].

Obtaining the linearisation of the non-linear forward maps (2.5) and (2.8) via arguments such as those given in Section 3.3 is a basic application of perturbative methods for inhomogeneous linear PDEs. The current proofs are taken from [94], [92], [96] where these problems where apparently first investigated in the setting of LAN expansions and information operators, with no claim of priority. Proposition 3.3.4 is taken from [96]. The techniques for the Schrödinger equation again extend to more complex inverse problems where a base differential operator $\mathscr{D}$ is perturbed by some potential, see, e.g., [92], [23] for the case where $\mathscr{D}$ is the geodesic vector field on the unit disk related to (1.1).

Chapter 4

Bernstein-von Mises theorems

A classical result of mathematical statistics due to Laplace, Bernstein, von Mises, Le Cam and van der Vaart gives general conditions under which posterior distributions $\Pi(\cdot|D_N)$ on parameter spaces Θ of fixed finite dimension are asymptotically approximated by a normal distribution centred at an efficient estimator (say the maximum likelihood estimator, or the posterior mean) and with 'inverse information covariance matrix' – see [83, 127]. This phenomenon occurs for general priors and likelihood models and can hence be understood as an expression of a 'universality' principle that posterior measures necessarily resemble the shape of Gaussians at least for large sample sizes. While this can be observed empirically in real data, the mathematical results substantiating it are called Bernstein-von Mises (BvM) theorems (to be distinguished from 'Laplace approximations' discussed in the next chapter). A main appeal of rigorous such theorems is that they provide objective statistical guarantees for posterior credible sets and hence for the Bayesian approach to *uncertainty quantification* (UQ) and the construction of 'error bars' for algorithmic outputs – see Section 4.1.3 for more on this.

While the BvM theorem is well understood in 'parametric' finite-dimensional models, obtaining high- or infinite-dimensional versions of it poses fundamental challenges. Already in basic conjugate Gaussian sequence space models examples for the invalidity of a Bernstein-von Mises approximation can be given, see [38, 54, 84] and also [72] for high-dimensional regression. However, following ideas from Le Cam theory and semi-parametric statistics, [28–30] developed techniques that allow to obtain BvM type results also in infinite-dimensional models, at least in 'weak enough topologies'. In this chapter we lay out some main ideas of this approach in the Gaussian regression model (1.11) and in the context of PDE inverse problems. Roughly speaking the goal is to describe conditions such that the 'cylindrical' laws of posterior measures characterised by all one-dimensional statistics $\langle \theta, \psi \rangle_{L^2}$ for a convergence-determining class of smooth test functions ψ are approximated by an infinite-dimensional Gaussian measure with a covariance structure that is information-theoretically optimal (in the sense of Theorem 3.1.4). [This convergence can be upgraded to be 'uniform' in ψ at the expense of technicalities that we wish to avoid here, but see after the proof of Theorem 4.2.1 for discussion.] The new analytical issue that arises is the inversion of the information operator appearing in the information equation (3.16). We will show that a semi-parametric BvM theorem can be proved when sufficiently regular solutions to the information equation exist. This in turn leads to new PDE questions of its own and we spell out what can happen for our two model examples with the Schrödinger and divergence form equations (2.5), (2.8).

4.1 Gaussian asymptotics for cylindrical laws

In this section we will show that if we can find sufficiently regular solutions to the 'information equation' from (3.16) for any given test function $\psi \in \Theta$, then as $N \to \infty$, a Gaussian approximation for the one-dimensional posterior statistics $\langle \theta, \psi \rangle_{L^2}$, $\theta \sim \Pi(\cdot|(Y_i, X_i)_{i=1}^N)$, holds true under some natural additional quantitative assumptions on the forward map $\mathscr{G}$ and the Gaussian process prior Π.

Colloquially one says that 'random via the data' probability measures $\mu_N \equiv \mu_N|(Y_i, X_i)_{i=1}^N$ *converge weakly in probability* to a limiting normal distribution μ. To make mathematical sense of this, one takes a metric d_{weak} for weak convergence of probability measures on the underlying space (see [47]) and shows the convergence to zero in $P_{\theta_0}^N$-probability of the real random variable $d_{weak}(\mu_N, \mu)$. For random variables $Z_N \sim \mu_N, Z \sim \mu$ we equivalently say that "$Z_N \to^d Z$ converges in distribution in $P_{\theta_0}^N$-probability" if

$$d_{weak}(\mu_N, \mu) \to^{P_{\theta_0}^N} 0 \text{ as } N \to \infty, \tag{4.1}$$

in the remainder of this section.

4.1.1 Asymptotic normality of linear functionals of the posterior

We consider again the general Gaussian regression model (1.11) with linear parameter space $\Theta \subset L^2_\zeta(\mathcal{Z}, W)$ where $\mathcal{Z}$ is a bounded domain in $\mathbb{R}^d$ with smooth boundary. We enforce on the forward map $\mathscr{G} : \Theta \to L^2_\lambda(\mathcal{X}, V)$ the Conditions 2.1.1 and 2.1.4 from earlier. We also suppose the linearisation Condition 3.1.1 holds for a linear operator $\mathbb{I}_{\theta_0}$ and with tangent space $H = \Theta$.

To simplify some proofs we slightly strengthen the forward regularity hypotheses to a L^∞-Lipschitz property.

Condition 4.1.1. *For all M and $\theta_1, \theta_2 \in \Theta$ such that $\|\theta_i\|_{\mathcal{R}} \leq M$ there exists a constant L such that*

$$\|\mathscr{G}(\theta_1) - \mathscr{G}(\theta_2)\|_\infty \leq L\|\theta_1 - \theta_2\|_\infty.$$

Moreover, the linear operator $\mathbb{I}_{\theta_0}$ is continuous from $(\Theta, \|\cdot\|_\infty)$ to $L^\infty(\mathcal{X}, V)$.

This condition can be verified along similar lines as in Propositions 2.1.2 and 2.1.3, involving a few extra technicalities. For $\mathscr{G}$ arising with the Schrödinger equation, see Ex. 4.3.1.

When considering inference on linear functionals $\langle \psi, \theta \rangle_{L^2(\mathcal{Z})}$ of θ, we require the existence of sufficiently regular solutions of the information equation (3.16) arising from the 'information' operator $\mathbb{I}^*_{\theta_0}\mathbb{I}_{\theta_0}$ from Definition 3.1.2 with $\Theta = H$. Formally we assume:

Condition 4.1.2. *Given $\psi \in \Theta$ suppose there exists $\bar{\psi} = \bar{\psi}_{\theta_0} \in \Theta$ such that $\mathbb{I}^*_{\theta_0}\mathbb{I}_{\theta_0}\bar{\psi} = \psi$, that is, $\langle \mathbb{I}^*_{\theta_0}\mathbb{I}_{\theta_0}\bar{\psi} - \psi, h\rangle_{L^2_\zeta(\mathcal{Z},W)} = 0$ for all $h \in \Theta$.*

As prior $\Pi = \Pi_N$ we take a (rescaled) Gaussian probability measure on Θ from Theorem 2.2.2. Next to the 'regularisation sets' Θ_N from that theorem, the proofs that follow require an initial posterior contraction rate $\bar{\delta}_N$ in $\|\cdot\|_\infty$-norm rather than just the L^2_ζ-norm. If the regularisation space $\mathcal{R}$ satisfies $\mathcal{R} \subseteq H^\beta$ for some $\beta > d/2$, then such a (not necessarily sharp) contraction rate can be obtained by interpolation and the Sobolev imbedding.

Proposition 4.1.3. *Suppose $\mathscr{G}, \Pi', \Pi_N, \mathcal{R}, \theta_0, \delta_N$ are as in Theorem 2.2.2 with regularisation space $\mathcal{R}$ continuously embedded into $H^\beta(\mathcal{Z})$ for some $\beta > d/2$. Suppose further that Condition 2.1.4 holds for such $\mathcal{R}$ and some $\eta > 0$. Define*

$$\Theta_{N,M,\infty} = \{\theta \in \Theta : \|\theta\|_{\mathcal{R}} \leq M, \|\theta - \theta_0\|_\infty \leq M\bar{\delta}_N\} \tag{4.2}$$

where M is a fixed constant and where

$$\bar{\delta}_N = \delta_N^{\eta(\beta-\beta')/\beta}, \quad \text{any } \beta' \in (d/2, \beta). \tag{4.3}$$

Then for all $b > 0$ we can choose M large enough such that as $N \to \infty$,

$$P^N_{\theta_0}\big(\Pi_N\big(\theta \in \Theta_{N,M,\infty}|D_N\big) \leq 1 - e^{-bN\delta_N^2}\big) \to 0. \tag{4.4}$$

Proof. By the Sobolev imbedding (6.4) and interpolation (6.5) we have

$$\|\theta - \theta_0\|_\infty \lesssim \|\theta - \theta_0\|_{H^\beta} \lesssim \|\theta - \theta_0\|^\vartheta_{L^2}\|\theta - \theta_0\|^{1-\vartheta}_{H^\beta} \tag{4.5}$$

where $\vartheta = (\beta - \beta')/\beta$. By hypothesis on $\mathcal{R}$ the norms $\|\theta - \theta_0\|_{H^\beta}$ are bounded by $M + \|\theta_0\|_{\mathcal{R}} < \infty$, so the limit (2.31) from Theorem 2.3.1 implies the desired result. □

The regularisation set $\Theta_{N,M,\infty}$ will play a quantitative role in the proofs to follow. It permits control of the non-linearity of the log-likelihood function (1.15), the discretisation errors arising from statistical sampling, and – as will be relevant – the sensitivity of Π_N with respect to small perturbations in $\bar{\psi}$-directions. To this end, let $\mathscr{J}_N$ be an upper bound for the following ('Dudley'-type) integral of the Kolmogorov metric entropy of $\Theta_{N,M,\infty}$;

$$\mathscr{J}_N(s,t) \geq \int_0^s \sqrt{\log 2N(\Theta_{N,M,\infty}, \|\cdot\|_\infty, t\epsilon)}d\epsilon, \quad s,t>0, \tag{4.6}$$

where $N(\Theta_{N,M,\infty}, \|\cdot\|_\infty, \epsilon)$ denotes the minimal number of $\|\cdot\|_\infty$-norm balls of radius ϵ required to cover the set $\Theta_{N,M,\infty}$.

Condition 4.1.4. *Assume $\sqrt{N}\delta_N^2 \to 0$ and that for $\bar{\delta}_N$ as in (4.3)*

$$\sqrt{N}\bar{\delta}_N^2 \mathscr{J}_N(1, \bar{\delta}_N^2) \to 0, \tag{4.7}$$

as $N \to \infty$. Further suppose that $\bar{\psi}_{\theta_0}$ from Condition 4.1.2 belongs to $\mathcal{H} \cap \mathcal{R}$ and let ρ_{θ_0} be the error term from Condition 3.1.1. For σ_N a sequence such that for all N large enough and all $t \in \mathbb{R}$ fixed,

$$\sigma_N \geq \sup_{\theta \in \Theta_{N,M,\infty}} \rho_{\theta_0}\Big[\theta - \theta_0 - \frac{t}{\sqrt{N}}\bar{\psi}_{\theta_0}\Big],$$

assume that as $N \to \infty$,

$$\max\Big(N(\sigma_N^2 + \sigma_N\bar{\delta}_N), \sqrt{N}\mathscr{J}_N(\sigma_N, 1), \bar{\delta}_N\sqrt{\log N}\,\mathscr{J}_N^2(\sigma_N, 1)/\sigma_N^2\Big) \to 0. \tag{4.8}$$

When the approximation in Condition 3.1.1 is quadratic ($\rho_{\theta_0}(h) = O(\|h\|_\infty^2)$), then (4.7) and (4.8) reduce to the much simpler conditions $N\bar{\delta}_N^3 \to 0, \beta > 2d$, in light of $\mathcal{R} \subset H^\beta(\mathcal{Z})$, see the proof of Theorem 4.2.1 below.

Our first theorem concerns an asymptotic approximation of the induced laws of the posterior distribution of $\langle \psi, \theta \rangle_{L^2}$ by a Gaussian distribution centred at

$$\hat{\Psi}_N = \langle \psi, \theta_0 \rangle_{L^2_\zeta(\mathcal{Z})} + \frac{1}{N}\sum_{i=1}^N \langle \mathbb{I}_{\theta_0}\bar{\psi}_{\theta_0}(X_i), \varepsilon_i \rangle_V. \tag{4.9}$$

We recall that the next limit is to be understood in the sense of (4.1).

Theorem 4.1.5. *Suppose Π_N, θ_0 are as in Proposition 4.1.3 and that the forward map $\mathscr{G}$ satisfies Conditions 2.1.1, 2.1.4, 3.1.1 with $H = \Theta$ and 4.1.1. Assume further that $\psi \in \Theta$ is such that Condition 4.1.2 holds for some $\bar{\psi}_{\theta_0}$ for which Condition 4.1.4 can be verified. Let $\theta \sim \Pi(\cdot|(Y_i, X_i)_{i=1}^N)$ be a posterior draw. Then we have as $N \to \infty$ and in $P_{\theta_0}^N$-probability,*

$$\sqrt{N}\big(\langle \theta, \psi \rangle_{L^2_\zeta(\mathcal{Z})} - \hat{\Psi}_N\big)|(Y_i, X_i)_{i=1}^N \to^d N(0, \|\mathbb{I}_{\theta_0}\bar{\psi}_{\theta_0}\|^2_{L^2_\lambda(\mathcal{X},V)}).$$

Proof. The plan is to prove convergence of the moment generating functions (Laplace transforms) of $\sqrt{N}\big(\langle\theta,\psi\rangle_{L^2(\mathcal{Z})} - \hat{\Psi}_N\big)|(Y_i,X_i)_{i=1}^N$ which implies weak convergence by standard arguments. We recall the notation $D_N := (Y_i,X_i)_{i=1}^N$ for the data vector. We will also write shorthand $\bar{\psi} = \bar{\psi}_{\theta_0}$ in this proof. The proof is split into four separate steps.

Localisation of the posterior measure

By Condition 4.1.4 the function $\bar{\psi}$ from Condition 4.1.2 defines an element of the RKHS $\mathcal{H}_N = \sqrt{N}\delta_N\mathcal{H}$ of Π_N with RKHS norm

$$\|\cdot\|_{\mathcal{H}_N} = \sqrt{N}\delta_N\|\cdot\|_{\mathcal{H}}.$$

If $\theta \sim \Pi_N$ then by definition of RKHS the random variable $\langle\theta,\bar{\psi}\rangle_{\mathcal{H}_N}$ has a $N(0,\|\bar{\psi}\|^2_{\mathcal{H}_N})$ distribution. Hence if we define

$$T_N = \Big\{\theta\in\Theta : \frac{|\langle\theta,\bar{\psi}\rangle_{\mathcal{H}_N}|}{\|\bar{\psi}\|_{\mathcal{H}_N}} > \sqrt{B}\sqrt{N}\delta_N\Big\},\ B>0,$$

the tail inequality for standard normal random variables implies that $\Pi(T_N) \leq e^{-BN\delta_N^2}$. So by Theorem 2.2.2 (and the remark at the beginning of its proof), and for $\Theta_{N,M,\infty}$ from Proposition 4.1.3, if we set

$$\bar{\Theta}_N := \Theta_{N,M,\infty}\cap T_N^c \quad \text{then} \quad \Pi(\bar{\Theta}_N^c|D_N) = O_{P^N_{\theta_0}}(e^{-bN\delta_N^2}) = o_{P^N_{\theta_0}}(1) \tag{4.10}$$

for any $b>0$ and as $N\to\infty$, as long as we choose M,B large enough. In the proofs that follow we consider $\theta\sim\Pi^{\bar{\Theta}_N}(\cdot|D_N)$ where the posterior (1.14) is taken to arise from prior probability measure

$$\Pi^{\bar{\Theta}_N} \equiv \frac{\Pi(\cdot\cap\bar{\Theta}_N)}{\Pi(\bar{\Theta}_N)}$$

equal to Π restricted to $\bar{\Theta}_N$ from (4.2) and renormalised. Standard arguments (Ex. 4.3.2) then imply, for $\|\cdot\|_{TV}$ the total variation distance on probability measures on Θ, that as $N\to\infty$

$$\|\Pi(\cdot|D_N) - \Pi^{\bar{\Theta}_N}(\cdot|D_N)\|_{TV} \leq 2\Pi(\bar{\Theta}_N^c|D_N) \to^{P^N_{\theta_0}} 0, \tag{4.11}$$

and as a consequence for any metric d_{weak} for weak convergence also

$$d_{weak}(\Pi(\cdot|D_N),\Pi^{\bar{\Theta}_N}(\cdot|D_N)) \to^{P^N_{\theta_0}} 0.$$

It hence suffices to prove Theorem 4.1.5 for $\Pi^{\bar{\Theta}_N}(\cdot|D_N)$ instead of $\Pi(\cdot|D_N)$.

Perturbation of the posterior Laplace transform

Proposition 4.1.6. *For $\theta, \psi \in \Theta$ and $\bar{\psi} = \bar{\psi}_{\theta_0}$ from Condition 4.1.2, define*

$$\theta_{(t)} = \theta - t\frac{\bar{\psi}_{\theta_0}}{\sqrt{N}},\ t \in \mathbb{R}.$$

Let $\hat{\Psi}_N$ be as in (4.9), $\bar{\Theta}_N$ as in (4.10), and ℓ_N as in (1.15). Then we have for every fixed $t \in \mathbb{R}$ and a sequence $R_N = o_{P_{\theta_0}^N}(1)$ as $N \to \infty$

$$E^{\Pi^{\bar{\Theta}_N}}\big[\exp\{t\sqrt{N}\big(\langle\theta,\psi\rangle_{L^2(\mathcal{Z})} - \hat{\Psi}_N\big)\}|D_N\big] = e^{\frac{t^2}{2}\|\mathbb{I}_{\theta_0}\bar{\psi}\|^2_{L^2(\mathcal{X})}} \times \frac{\int_{\bar{\Theta}_N} e^{\ell_N(\theta_{(t)})}d\Pi(\theta)}{\int_{\bar{\Theta}_N} e^{\ell_N(\theta)}d\Pi(\theta)} \times e^{R_N}.$$

Proof. For W_N as in (3.8) with $h = \bar{\psi}$, the posterior Laplace transform equals

$$E^{\Pi^{\bar{\Theta}_N}}\big[e^{t\sqrt{N}(\langle\theta,\psi\rangle_{L^2(\mathcal{Z})} - \hat{\Psi}_N)}|D_N\big] = \frac{\int_{\bar{\Theta}_N} e^{t\sqrt{N}\langle\theta-\theta_0,\psi\rangle_{L^2(\mathcal{Z})} - tW_N + \ell_N(\theta) - \ell_N(\theta_{(t)}) + \ell_N(\theta_{(t)})}d\Pi(\theta)}{\int_{\bar{\Theta}_N} e^{\ell_N(\theta)}d\Pi(\theta)}$$

The main step in the proof is a uniform in $\theta \in \bar{\Theta}_N$ perturbation expansion of the log-likelihood ratios under $P_{\theta_0}^N$. We can write

$$\begin{aligned}
&\ell_N(\theta) - \ell_N(\theta_{(t)})\\
&= -\frac{1}{2}\sum_{i=1}^N \big(|Y_i - \mathscr{G}(\theta)(X_i)|_V^2 - |Y_i - \mathscr{G}(\theta_{(t)})(X_i)|_V^2\big)\\
&= -\frac{1}{2}\sum_{i=1}^N \big(|\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta)(X_i) + \varepsilon_i|_V^2 - |\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta_{(t)})(X_i) + \varepsilon_i|_V^2\big)\\
&= -\sum_{i=1}^N \Big(\langle\varepsilon_i, \mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta)(X_i)\rangle_V - \langle\varepsilon_i, \mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta_{(t)})(X_i)\rangle_V\Big)\\
&\qquad - \frac{1}{2}\sum_{i=1}^N \Big(|\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta)(X_i)|_V^2 - |\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta_{(t)})(X_i)|_V^2\Big) \equiv I + II.
\end{aligned}$$

About term I, we 'linearise' the map $\mathscr{G}$ at θ_0 in each inner product to obtain

$$\begin{aligned}
I &= \sum_{i=1}^N \langle\varepsilon_i, D\mathscr{G}_{\theta_0}(X_i)[\theta - \theta_{(t)}]\rangle_V\\
&\quad + \sum_{i=1}^N \langle\varepsilon_i, \mathscr{G}(\theta)(X_i) - \mathscr{G}(\theta_0)(X_i) - D\mathscr{G}_{\theta_0}(X_i)[\theta - \theta_0]\rangle_V\\
&\quad - \sum_{i=1}^N \langle\varepsilon_i, \mathscr{G}(\theta_{(t)})(X_i) - \mathscr{G}(\theta_0)(X_i) - D\mathscr{G}_{\theta_0}(X_i)[\theta_{(t)} - \theta_0]\rangle_V
\end{aligned}$$

which we write as

$$\frac{t}{\sqrt{N}}\sum_{i=1}^{N}\langle \varepsilon_i, D\mathscr{G}_{\theta_0}(X_i)[\bar{\psi}]\rangle_V + R_{(0)}(\theta) - R_{(t)}(\theta) = tW_N + R_{(0)}(\theta) - R_{(t)}(\theta),$$

noting that $\theta_{(0)} = \theta$ and where the 'remainder empirical processes' are given by

$$R_{(t)} \equiv \sum_{i=1}^{N}\langle \varepsilon_i, \mathscr{G}(\theta_{(t)})(X_i) - \mathscr{G}(\theta_0)(X_i) - D\mathscr{G}_{\theta_0}(X_i)[\theta_{(t)} - \theta_0]\rangle_V.$$

We show in Lemma 4.1.7 below that for all $t \in \mathbb{R}$ fixed,

$$\sup_{\theta \in \Theta_{N,M,\infty}} |R_{(t)}(\theta)| = o_{P^N_{\theta_0}}(1) \tag{4.12}$$

so that these terms form a part of the sequence R_N.

For term II, with E_λ denoting expectation under the X_i's only, we have

$$\begin{aligned}
&-\frac{1}{2}\sum_{i=1}^{N}\Big(|\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta)(X_i)|_V^2 - E_\lambda|\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta)(X_i)|_V^2 \\
&\qquad - |\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta_{(t)})(X_i)|_V^2 + E_\lambda|\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta_{(t)})(X_i)|_V^2\Big) \\
&-\frac{N}{2}\|\mathscr{G}(\theta_0) - \mathscr{G}(\theta)\|_{L^2(\mathcal{X})}^2 + \frac{N}{2}\|\mathscr{G}(\theta_0) - \mathscr{G}(\theta_{(t)})\|_{L^2(\mathcal{X})}^2.
\end{aligned}$$

The sums in the first two lines are empirical processes and are shown in Lemma 4.1.8 below to be $o_{P^N_{\theta_0}}(1)$ uniformly in $\theta \in \Theta_{N,M,\infty}$ for every fixed t, and can thus also be absorbed into R_N.

For the terms in the last line of the last display, we can further decompose

$$\begin{aligned}
\|\mathscr{G}(\theta_0) - \mathscr{G}(\theta_{(t)})\|_{L^2(\mathcal{X})}^2 &= \|\mathscr{G}(\theta_{(t)}) - \mathscr{G}(\theta_0) - D\mathscr{G}_{\theta_0}[\theta_{(t)} - \theta_0] + D\mathscr{G}_{\theta_0}[\theta_{(t)} - \theta_0]\|_{L^2(\mathcal{X})}^2 \\
&= \|D\mathscr{G}_{\theta_0}[\theta_{(t)} - \theta_0]\|_{L^2(\mathcal{X})}^2 \\
&\quad + 2\langle D\mathscr{G}_{\theta_0}[\theta_{(t)} - \theta_0], \mathscr{G}(\theta_{(t)}) - \mathscr{G}(\theta_0) - D\mathscr{G}_{\theta_0}[\theta_{(t)} - \theta_0]\rangle_{L^2(\mathcal{X})} \\
&\quad + \|\mathscr{G}(\theta_{(t)}) - \mathscr{G}(\theta_0) - D\mathscr{G}_{\theta_0}[\theta_{(t)} - \theta_0]\|_{L^2(\mathcal{X})}^2
\end{aligned}$$

including also the case $\theta = \theta_{(0)}$ by convention for $t = 0$. Now using Conditions 3.1.1, 4.1.4 and the Cauchy-Schwarz inequality the last two remainder terms are bounded by a constant multiple of

$$\sup_{\theta \in \Theta_{N,M,\infty}} \left[\rho_{\theta_0}^2(\theta_{(t)} - \theta_0) + \|\theta_{(t)} - \theta_0\|_{L^2}\rho_{\theta_0}(\theta_{(t)} - \theta_0)\right] \lesssim \sigma_N^2 + \sigma_N\bar{\delta}_N = o(1/N).$$

The remaining terms in the expansion are

$$\begin{aligned}
&\frac{N}{2}\big(\big\|D\mathscr{G}_{\theta_0}[\theta-\theta_0-\frac{t}{\sqrt{N}}\bar{\psi}]\big\|^2_{L^2(\mathcal{X},V)}-\big\|D\mathscr{G}_{\theta_0}[\theta-\theta_0]\big\|^2_{L^2(\mathcal{X},V)}\big)\\
&=-t\sqrt{N}\langle D\mathscr{G}_{\theta_0}[\theta-\theta_0],D\mathscr{G}_{\theta_0}[\bar{\psi}]\rangle_{L^2(\mathcal{X},V)}+\frac{t^2}{2}\|D\mathscr{G}_{\theta_0}[\bar{\psi}]\big\|^2_{L^2(\mathcal{X},V)}\\
&=-t\sqrt{N}\langle\theta-\theta_0,\mathbb{I}^*_{\theta_0}\mathbb{I}_{\theta_0}\bar{\psi}\rangle_{L^2(\mathcal{Z},W)}+\frac{t^2}{2}\|\mathbb{I}_{\theta_0}[\bar{\psi}]\big\|^2_{L^2(\mathcal{X},V)}
\end{aligned}$$

which, combined with Condition 4.1.2, the bounds from term I and the identity in the first display in this proof, implies the result. □

Stochastic bounds on remainder terms and discretisation error

The following two lemmas use tools from infinite-dimensional probability to bound the collections of empirical processes appearing as remainder terms in the proof of Proposition 4.1.6. While that proposition considers localisation to the sets $\bar{\Theta}_N$, the following bounds actually hold uniformly in the larger classes $\Theta_{N,M,\infty}$ from (4.2).

Lemma 4.1.7. *We have (4.12).*

Proof. For t fixed define new functions $g_\theta:\mathcal{X}\to V$ as

$$g_\theta=\mathscr{G}(\theta_{(t)})(\cdot)-\mathscr{G}(\theta_0)(\cdot)-D\mathscr{G}_{\theta_0}(\cdot)[\theta_{(t)}-\theta_0].$$

Then the remainder term from (4.12), viewed as a stochastic process indexed by $\theta\in\Theta_{N,M,\infty}$, equals a centred (since $E\varepsilon_i=0$) *empirical process* for the jointly i.i.d. variables (X_i,ε_i) of the form

$$|R_{(t)}(\theta)|\equiv\big|\sum_{i=1}^N\langle\varepsilon_i,g_\theta(X_i)\rangle_V\big|\equiv\big|\sum_{j=1}^{p_V}\sum_{i=1}^N\varepsilon_{i,j}g_{\theta,j}(X_i)\big|\le\sum_{j=1}^{p_V}\big|\sum_{i=1}^N\varepsilon_{i,j}g_{\theta,j}(X_i)\big|.$$

Here $g_{\theta,j}$ are the entries of the vector field $g_\theta\in V$, and the $\varepsilon_{i,j}$ are all i.i.d. $N(0,1)$ variables. We will now bound the supremum over $\Theta_{N,M,\infty}$ of the each of the last p_V summands by using a moment inequality for the empirical process $\{\sum_{i=1}^N f_\theta(Z_i):f\in\mathcal{F}\}$ where, for every $1\le j\le p$ fixed (and with e denoting a real variable in this proof in slight abuse of notation),

$$f_\theta\in\mathcal{F}\equiv\mathcal{F}_j=\{f_\theta(z)=eg_{\theta,j}(x):\theta\in\Theta_{N,M,\infty}\},\quad z=(e,x)\in\mathbb{R}\times\mathcal{X},$$

and $Z_1,\dots,Z_N$ are i.i.d. copies of the variables $Z=(\varepsilon,X)\sim N(0,1)\times\lambda=P$.

We will apply Theorem 3.5.4 in [61] but to do so need to calculate some preliminary bounds: First, by independence of X, ε, the 'weak' variances of $\mathcal{F}$ are of order

$$\sup_{\theta \in \Theta_{N,M,\infty}} Ef_\theta^2(Z) = \sup_{\theta \in \Theta_{N,M,\infty}} Eg_{\theta,j}^2(X) \le \sup_{\theta \in \Theta_{N,M,\infty}} \rho_{\theta_0}^2(\theta_{(t)} - \theta_0) \le \sigma_N^2$$

by Conditions 3.1.1 and 4.1.4. Next, by Condition 4.1.1 and the definition of $\Theta_{N,M,\infty}$ we have

$$\sup_{\theta \in \Theta_{N,M,\infty}} \|g_{\theta,j}\|_\infty \lesssim \|\theta_{(t)} - \theta_0\|_\infty \lesssim \bar\delta_N(1 + \|\bar\psi\|_\infty) \lesssim \bar\delta_N.$$

As a consequence the preceding empirical process has point-wise envelopes

$$\sup_{\theta \in \Theta_{N,M,\infty}} |f_\theta(e,x)| \lesssim |e|\bar\delta_N \equiv F_N(e,x) \quad \forall (e,x) \in \mathbb{R} \times \mathcal{X},$$

in particular $F_N > 0$ P-a.s. and

$$\|F_N\|_{L^2(P)}^2 := \int_{\mathbb{R}\times\mathcal{X}} F_N^2(z)dP(z) \lesssim \bar\delta_N^2, \quad \|F_N\|_{L^2(Q)}^2 := \int_{\mathbb{R}\times\mathcal{X}} F_N^2(z)dQ(z) \simeq \bar\delta_N^2 s_Q^2,$$

where, for any (discrete, finitely supported) probability measure Q on $\mathbb{R} \times \mathcal{X}$, we have set $s_Q^2 := \int_{\mathbb{R}\times\mathcal{X}} e^2 dQ(e,x)$. Finally, we have again from Condition 4.1.1, for any $\theta, \theta' \in \Theta$ and some fixed constant c_0 that

$$\begin{aligned}
\|f_\theta - f_{\theta'}\|_{L^2(Q)} &:= \sqrt{\int_{\mathbb{R}}\int_{\mathcal{X}} e^2(g_{\theta,j}(x) - g_{\theta',j}(x))^2 dQ(e,x)} \\
&\le s_Q\|g_{\theta,j} - g_{\theta',j}\|_\infty \\
&\le s_Q(\|\mathscr{G}(\theta_{(t)}) - \mathscr{G}(\theta'_{(t)})\|_\infty + \|\mathbb{I}_{\theta_0}[\theta_{(t)} - \theta'_{(t)}]\|_\infty) \\
&\le c_0\|F_N\|_{L^2(Q)}\|\theta - \theta'\|_\infty/\bar\delta_N.
\end{aligned}$$

We conclude that any $\bar\delta_N\epsilon/c_0$-covering of $\Theta_{N,M,\infty}$ for the norm $\|\cdot\|_\infty$ induces a $\|F_N\|_{L^2(Q)}\epsilon$-covering of $\mathcal{F}$ for the $L^2(Q)$ norm, and so $J(\mathcal{F}, F, s)$ in (3.169) in [61] is bounded by a constant multiple of our $\mathscr{J}_N(s, \bar\delta_N)$ (using also Lemma 3.5.3a in [61]). With these preparations, we can now apply Theorem 3.5.4 in [61] where for our choice of envelope F_N we can take $\|U\|_{L^2(P)}$ in that theorem bounded by a constant multiple of $\sqrt{\log N}\bar\delta_N$ (using independence of X, ε and also Lemma 2.3.3 in [61]). The upper bound (3.171) in [61] then implies that

$$E \sup_{\theta \in \Theta_{N,M,\infty}} \Big|\sum_{i=1}^N f_\theta(Z_i)\Big| \lesssim \sqrt{N} \max\left[\bar\delta_N \mathscr{J}_N(\sigma_N/\bar\delta_N, \bar\delta_N), \frac{\sqrt{\log N}\bar\delta_N^3 \mathscr{J}_N^2(\sigma_N/\bar\delta_N, \bar\delta_N)}{\sqrt{N}\sigma_N^2}\right]$$

which in turn, using the substitution $\bar{\delta}_N\epsilon = \rho$ in (4.6), is bounded by a constant multiple of the maximum of the second and third terms appearing in (4.8). Hence the remainder terms from (4.12) converge to zero in expectation, and then also in probability (by Markov's inequality). [Let us finally note that, strictly speaking, the application of Theorem 3.5.4 in [61] requires $0 \in \mathcal{F}$ and $\mathcal{F}$ countable: If $\|\theta_0\|_{\mathcal{R}} < M$ then $g_\theta = 0$ for $\theta = \theta_0 - (t/\sqrt{N})\bar{\psi} \in \Theta_{N,M,\infty}$ and N large enough, so $0 \in \mathcal{F}$. Otherwise we can recenter f_θ at f_{θ_*} for some arbitrary θ_* and use a standard (one-dimensional) moment bound for $E|\sum_{i=1}^N f_{\theta_*}(Z_i)| \leq \sqrt{N}\sigma_N \to 0$. One then applies the previous argument to the class $\mathcal{F} - f_{\theta_*}$, so that the same overall bound holds true also in this case. Finally, by continuity of $\theta \mapsto g_{\theta,j}$ on the totally bounded set $\Theta_{N,M,\infty}$, the supremum of the empirical process can be realised over a countable dense subset of $\Theta_{N,M,\infty}$, so the assumption that $\mathcal{F}$ be countable can be met, too.] $\square$

Lemma 4.1.8. *We have for any $t \in \mathbb{R}$ that*

$$\sup_{\theta \in \Theta_{N,M,\infty}} \Big| \sum_{i=1}^N \Big(|\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta_{(t)})(X_i)|_V^2 - E_\lambda |\mathscr{G}(\theta_0)(X_i) - \mathscr{G}(\theta)(X_i)|_V^2 \Big) \Big| = o_{P_{\theta_0}^N}(1)$$

Proof. We will obtain a bound for the supremum of the empirical process $\{\sum_{i=1}^N (f(X_i) - Ef(X_i)) : f \in \mathcal{F}\}$, this time with indexing class

$$\mathcal{F} = \{f_\theta = |\mathscr{G}(\theta_0)(\cdot) - \mathscr{G}(\theta_{(t)})(\cdot)|_V^2 : \theta \in \Theta_{N,M,\infty}\}.$$

Using Condition 4.1.1, the envelopes of $\mathcal{F}$ can be taken to be

$$\sup_{\theta \in \Theta_{N,M,\infty}} \|\mathscr{G}(\theta_0) - \mathscr{G}(\theta_{(t)})\|_\infty^2 \lesssim \sup_{\theta \in \Theta_{N,M,\infty}} \|\theta_0 - \theta_{(t)}\|_\infty^2 \lesssim \bar{\delta}_N^2 \equiv F,$$

and we also have, since $\|\mathscr{G}(\theta)\|_\infty \leq U$ for $\theta \in \Theta_{N,M,\infty}$ by Condition 2.1.1, that

$$\|f_\theta - f_{\theta'}\|_\infty \lesssim \|\theta - \theta'\|_\infty \quad \forall \theta, \theta' \in \Theta_{N,M,\infty}.$$

This implies, similar to the proof in the previous lemma, that a $c_0\bar{\delta}_N^2\epsilon$-covering of $\Theta_{N,M,\infty}$ for the $\|\cdot\|_\infty$-norm (and c_0 a small but fixed constant) induces a $\|F\|_{L^2(Q)}\epsilon$-covering of $\mathcal{F}$ for the $L^2(Q)$-norm (Q any probability measure), and that the functional $J(\mathcal{F}, F, s)$ in (3.169) in [61] is bounded by a constant multiple of our $\mathscr{J}(s, \bar{\delta}_N^2)$. The convergence to zero required in the lemma now follows from Theorem 3.5.4 in [61], in fact Remark 3.5.5 after it, the requirement (4.7) from Condition 4.1.4, and Markov's inequality. $\square$

Gaussian change of variables

We now control the ratio of Gaussian integrals appearing in Proposition 4.1.6.

Proposition 4.1.9. *As $N \to \infty$ we have for any fixed $t \in \mathbb{R}$ that*

$$\frac{\int_{\bar{\Theta}_N} e^{\ell_N(\theta_{(t)})} d\Pi(\theta)}{\int_{\bar{\Theta}_N} e^{\ell_N(\theta)} d\Pi(\theta)} \to^{P_{\theta_0}^N} 1.$$

Proof. If we denote by Π_t the Gaussian law of $\theta_{(t)} = \theta - (t/\sqrt{N})\bar{\psi}$, then the Cameron-Martin theorem (e.g., Theorem 2.6.13 in [61]) provides the formula for the Radon-Nikodym density of

$$\frac{d\Pi_t}{d\Pi}(\theta) = \exp\left\{\frac{t}{\sqrt{N}}\langle \theta, \bar{\psi}\rangle_{\mathcal{H}_N} - \frac{t^2}{2N}\|\bar{\psi}\|^2_{\mathcal{H}_N}\right\}, \quad \theta \sim \Pi, \quad \bar{\psi} \in \mathcal{H}_N.$$

The ratio in the proposition thus equals

$$\frac{e^{-\frac{t^2}{2N}\|\bar{\psi}\|^2_{\mathcal{H}_N}} \int_{\bar{\Theta}_{N,t}} e^{\ell_N(\vartheta)} e^{\frac{t}{\sqrt{N}}\langle \vartheta, \bar{\psi}\rangle_{\mathcal{H}_N}} d\Pi(\vartheta)}{\int_{\bar{\Theta}_N} e^{\ell_N(\theta)} d\Pi(\theta)}, \quad \text{where} \quad \bar{\Theta}_{N,t} = \{\vartheta = \theta_{(t)} : \theta \in \bar{\Theta}_N\}.$$

Uniformly in $\theta \in T_N^c \subset \bar{\Theta}_N$ from (4.10) we have as $N \to \infty$ that

$$\frac{|t|}{\sqrt{N}}|\langle \theta, \bar{\psi}\rangle_{\mathcal{H}_N}| \lesssim \delta_N \|\bar{\psi}\|_{\mathcal{H}_N} \lesssim \sqrt{N}\delta_N^2 \|\bar{\psi}\|_{\mathcal{H}} \to 0$$

by Condition 4.1.4, which also implies that $(t^2/N)\|\bar{\psi}\|^2_{\mathcal{H}_N} = o(1)$ since $1/\sqrt{N} = o(\delta_N)$. Now since

$$\frac{|t|}{\sqrt{N}} \sup_{\vartheta \in \bar{\Theta}_{N,t}} |\langle \vartheta, \bar{\psi}\rangle_{\mathcal{H}_N}| \le \frac{|t|}{\sqrt{N}} \sup_{\theta \in T_N^c} |\langle \theta, \bar{\psi}\rangle_{\mathcal{H}_N}| + \frac{t^2}{N}\|\bar{\psi}\|^2_{\mathcal{H}_N}$$

we deduce from what precedes that the last ratio of integrals equals

$$e^{o(1)} \times \frac{\int_{\bar{\Theta}_{N,t}} e^{\ell_N(\vartheta)} d\Pi(\vartheta)}{\int_{\bar{\Theta}_N} e^{\ell_N(\theta)} d\Pi(\theta)} = e^{o(1)} \times \frac{\Pi(\bar{\Theta}_{N,t}|D_N)}{\Pi(\bar{\Theta}_N|D_N)}.$$

The denominator converges to 1 in $P_{\theta_0}^N$-probability by (4.10), and so does then the numerator, using again (4.10) and that $t\|\bar{\psi}\|_\infty/\sqrt{N} = o(\bar{\delta}_N)$ and $t\|\bar{\psi}\|_{\mathcal{R}}/\sqrt{N} = o(1)$ under the maintained assumptions. $\square$

Combining Propositions 4.1.6 and 4.1.9 we have for all $t \in \mathbb{R}$, as $N \to \infty$,

$$E^{\Pi^{\bar\Theta_N}}\left[\exp\{t\sqrt{N}\big(\langle\theta,\psi\rangle_{L^2_\zeta(\mathcal{Z})} - \hat\Psi_N\big)\}|D_N\right] \to \exp\left\{\frac{t^2}{2}\|\mathbb{I}_{\theta_0}\bar\psi\|^2_{L^2_\lambda(\mathcal{X})}\right\} \tag{4.13}$$

in $P^N_{\theta_0}$-probability, and therefore, using also (4.11), for $\theta \sim \Pi(\cdot|D_N)$,

$$\sqrt{N}\big(\langle\theta,\psi\rangle_{L^2_\zeta(\mathcal{Z})} - \hat\Psi_N\big)|D_N \to^d N(0, \|\mathbb{I}_{\theta_0}\bar\psi\|^2_{L^2_\lambda(\mathcal{X})}) \tag{4.14}$$

by the in $P^N_{\theta_0}$-probability version of the usual implication that convergence of Laplace transforms implies convergence in distribution (see Ex. 4.3.3). This completes the proof of Theorem 4.1.5. □

4.1.2 Asymptotic distribution of the posterior mean

To use an approximation as the one from Theorem 4.1.5 for applications to uncertainty quantification (see Section 4.1.3), we need to choose a *feasibly computable centring statistic* instead of the (infeasible) $\hat\Psi_N$. A desirable choice, both for inference and computation via MCMC, is

$$\langle\bar\theta_N,\psi\rangle_{L^2_\zeta(\mathcal{Z},W)}, \quad \text{where } \ \bar\theta_N = E^\Pi[\theta|(Y_i,X_i)_{i=1}^N]$$

is the mean of the posterior distribution. Under the maintained hypotheses, the Bochner integral $E^\Pi[\theta|(Y_i,X_i)_{i=1}^N]$ can be shown to exists for any given data vector $(Y_i,X_i)_{i=1}^N$.

Theorem 4.1.10. *In the setting of Theorem 4.1.5, if $\bar\theta_N = E^\Pi[\theta|(Y_i,X_i)_{i=1}^N]$ denotes the posterior mean, then we have as $N \to \infty$,*

$$\sqrt{N}\langle\theta-\bar\theta_N,\psi\rangle_{L^2_\zeta(\mathcal{Z})}|(Y_i,X_i)_{i=1}^N \to^d N(0,\|\mathbb{I}_{\theta_0}\bar\psi_{\theta_0}\|^2_{L^2_\lambda(\mathcal{X},V)}) \quad \text{in } P^N_{\theta_0} - \text{ probability.}$$

Moreover, as $N \to \infty$, we also have

$$\sqrt{N}\langle\bar\theta_N-\theta_0,\psi\rangle_{L^2_\zeta(\mathcal{Z},W)} \to^d N(0,\|\mathbb{I}_{\theta_0}\bar\psi_{\theta_0}\|^2_{L^2_\lambda(\mathcal{X},V)}).$$

The proof consists of a 'quantitative' uniform integrability argument employing the following lemma which provides a stochastic bound on the posterior second moments.

Lemma 4.1.11. *Under the hypotheses of Theorem 4.1.10 we have*

$$NE^\Pi\left[(\langle\theta,\psi\rangle_{L^2_\zeta(\mathcal{Z})} - \hat\Psi_N)^2|D_N\right] = O_{P^N_{\theta_0}}(1)$$

Proof. The left hand side in the last display is bounded by

$$2NE^{\Pi}\big[\langle\theta-\theta_0,\psi\rangle^2_{L^2_\zeta(\mathcal{Z})}|D_N\big]+2N(\hat{\Psi}_N-\langle\theta_0,\psi\rangle_{L^2_\zeta(\mathcal{Z})})^2$$

and in view of (4.9), the second term in the last decomposition is bounded in $P^N_{\theta_0}$-probability by the central limit theorem applied to W_N from (3.8) with $h=\psi_{\theta_0}$ (one also applies the continuous mapping theorem for $x\mapsto x^2$ and Prohorov's theorem ([127]) to deduce from convergence in distribution of NW_N^2 that it is uniformly tight.)

It hence remains to bound the first term in the last decomposition. Define $C_N=\{\|\theta-\theta_0\|_\infty\le M\bar{\delta}_N\}\subset\Theta$ for M to be chosen, and write the first quantity in the last display as (two times)

$$NE^{\Pi}\big[\langle\theta-\theta_0,\psi\rangle^2_{L^2_\zeta(\mathcal{Z})}1_{C_N}|D_N\big]+NE^{\Pi}\big[\langle\theta-\theta_0,\psi\rangle^2_{L^2_\zeta(\mathcal{Z})}1_{C_N^c}|D_N\big]=I+II. \quad (4.15)$$

To deal with term II, we apply the Cauchy-Schwarz inequality to obtain the bound

$$N\sqrt{E^{\Pi}\big[\langle\theta-\theta_0,\psi\rangle^4_{L^2(\mathcal{Z})}|D_N\big]}\sqrt{\Pi(\|\theta-\theta_0\|_\infty>M\bar{\delta}_N|D_N)}$$

and we now show that this term is bounded in $P^N_{\theta_0}$-probability: Using Lemma 1.3.3 and the arguments from step ii) in the proof of Theorem 2.2.2 to bound the probability of the sets A_N from (1.37), Markov's inequality and $E^N_{\theta_0}e^{\ell_N(\theta)-\ell_N(\theta_0)}=1$, as well as Proposition 4.1.3 with M large enough so that $b>A+3$, we obtain

$$\begin{aligned}
&P^N_{\theta_0}\Big(E^{\Pi}\big[\langle\theta-\theta_0,\psi\rangle^4|D_N]\Pi(\|\theta-\theta_0\|_\infty>M\bar{\delta}_N|D_N)>N^{-2}\Big)\\
&\le P^N_{\theta_0}\Big(E^{\Pi}\big[\langle\theta-\theta_0,\psi\rangle^4|D_N]e^{-bN\delta_N^2}>N^{-2}\Big)+o(1)\\
&\le P^N_{\theta_0}\left(\frac{\int_\Theta\langle\theta-\theta_0,\psi\rangle^4e^{\ell_N(\theta)-\ell_N(\theta_0)}d\Pi(\theta)}{\int_\Theta e^{\ell_N(\theta)-\ell_N(\theta_0)}d\Pi(\theta)}>e^{bN\delta_N^2}N^{-2},A_N\right)+o(1)\\
&\le\|\psi\|^4_{L^2(\mathcal{Z})}e^{(A+2-b)N\delta_N^2}N^2\int_\Theta\|\theta-\theta_0\|^4_{L^2(\mathcal{Z})}E^N_{\theta_0}e^{\ell_N(\theta)-\ell_N(\theta_0)}d\Pi(\theta)+o(1)\\
&\lesssim N^2e^{-N\delta_N^2}+o(1)\to 0
\end{aligned}$$

as $N\to\infty$, by hypothesis on δ_N,Π_N. Collecting what precedes implies that the term II in (4.15) is indeed $O_{P^N_{\theta_0}}(1)$.

The next step is to bound the term I in (4.15). Recalling that $\Pi^{\bar{\Theta}_N}[\cdot|D_N]$ denotes the posterior distribution arising from prior restricted and renormalised to $\bar{\Theta}_N$, we decompose

$$\begin{aligned}
NE^{\Pi}\big[\langle\theta-\theta_0,\psi\rangle^2_{L^2(\mathcal{Z})}1_{C_N}|D_N\big]&=NE^{\Pi^{\bar{\Theta}_N}}\big[\langle\theta-\theta_0,\psi\rangle^2_{L^2(\mathcal{Z})}1_{C_N}|D_N\big]\\
&+NE^{\Pi}\big[\langle\theta-\theta_0,\psi\rangle^2_{L^2(\mathcal{Z})}1_{C_N}|D_N\big]-NE^{\Pi^{\bar{\Theta}_N}}\big[\langle\theta-\theta_0,\psi\rangle^2_{L^2(\mathcal{Z})}1_{C_N}|D_N\big]=A+B.
\end{aligned}$$

For term A, using $x^2 \leq 2e^x, x \geq 0$, the definition of $\hat{\Psi}_N$ from (4.9) and $W_N = O_{P^N_{\theta_0}}(1)$ with $h = \bar{\psi}_{\theta_0}$ from (3.8), the limit (4.13) at $t = 1$ implies that for all N large enough and some $r_N = o_{P^N_{\theta_0}}(1)$,

$$A \leq 2e^{W_N + r_N} e^{\frac{1}{2}\|\mathbb{I}_{\theta_0}\bar{\psi}\|^2_{L^2(\mathcal{X},V)}},$$

and hence this term is stochastically bounded.

Finally, by definition of the events C_N, the term $|B|$ can be written as

$$\begin{aligned} &N\Big|\int_{C_N} \langle \theta - \theta_0, \psi\rangle^2_{L^2(\mathcal{Z})}[d\Pi(\theta|D_N) - d\Pi^{\bar{\Theta}_N}(\theta|D_N)]\Big| \\ &\leq MN\bar{\delta}_N^2\|\psi\|^2_{L^1(\mathcal{Z})}\|\Pi(\cdot|D_N) - \Pi^{\bar{\Theta}_N}(\cdot|D_N)\|_{TV} \\ &\lesssim N\bar{\delta}_N^2\Pi(\bar{\Theta}_N^c|D_N) \lesssim N\bar{\delta}_N^2 O_{P^N_{\theta_0}}(e^{-bN\delta_N^2}) = o_{P^N_{\theta_0}}(1), \end{aligned}$$

where we have used (4.11) and (4.10), completing the proof of the lemma. □

Now to prove the theorem note that by (4.14) and (4.1) we have for

$$Z_n|D_N \equiv \sqrt{N}(\langle\theta,\psi\rangle_{L^2_\zeta(\mathcal{Z})} - \hat{\Psi}_N)|D_N, \quad Z \sim N(0, \|\mathbb{I}_{\theta_0}\bar{\psi}\|^2_{L^2_\lambda(\mathcal{X})})$$

and d_{weak} any metric for weak convergence of laws $\mathcal{L}(\cdot)$ on $\mathbb{R}$,

$$d_{weak}\left(\mathcal{L}\big(Z_N|D_N\big), \mathcal{L}(Z)\big)\right) \to^{P^N_{\theta_0}}_{N\to\infty} 0. \tag{4.16}$$

The idea of the proof to follow is that the previous lemma implies (by uniform integrability) convergence of moments in the last limit (4.16), and thus that, since $EZ = 0$, the posterior mean equals $\hat{\Psi}_N$ up to a stochastic term of order $o(1/\sqrt{N})$. However, as the probability measures $\mathcal{L}(Z_N|D_N)$ to which this argument is applied are *random* via the data D_N, the proof requires some care. We will employ a contradiction argument: To prove Theorem 4.1.10, it suffices by Theorem 4.1.5, Slutsky's lemma and (3.8) with $h = \bar{\psi}_{\theta_0}$ to prove that as $N \to \infty$,

$$\sqrt{N}\big(\langle E^\Pi[\theta|D_N], \psi\rangle_{L^2_\zeta(\mathcal{Z})} - \hat{\Psi}_N\big) \to^{\Pr} 0, \tag{4.17}$$

where we write Pr for the probability measure $P^{\mathbb{N}}_{\theta_0}$ on the underlying measurable space $(\Omega, \mathcal{S}) := ((V \times \mathcal{X})^{\mathbb{N}}, \mathcal{S})$ supporting all data variables $(D_N, N \in \mathbb{N})$. Suppose the last limit does not hold true. Then there exists $\Omega' \in \mathcal{S}$ of positive probability $\Pr(\Omega') > \tau$ and $\zeta' > 0$ such that along a subsequence of N (still denoted by N) we have

$$|\sqrt{N}\big(\langle E^\Pi[\theta|D_N(\omega)], \psi\rangle_{L^2_\zeta(\mathcal{Z})} - \hat{\Psi}_N(\omega)\big)| \geq \zeta' > 0 \quad \text{for } \omega \in \Omega'. \tag{4.18}$$

Now since convergence in Pr-probability implies Pr-almost sure convergence along a subsequence, we can extract a further subsequence of N such that (4.16) holds almost surely, that is, on an event $\Omega_0 \subset \Omega$ such that $\Pr(\Omega_0) = 1$. For each fixed $\omega \in \Omega_0$ we can use the Skorohod imbedding (Theorem 11.7.2 in [47]) to construct (if necessary on a new probability space) new real random variables $\tilde{Z}_N, \tilde{Z}$ such that their laws satisfy

$$\mathcal{L}(\tilde{Z}_N) = \mathcal{L}\big(Z_N|D_N(\omega)\big), \quad \mathcal{L}(\tilde{Z}) = \mathcal{L}(Z), \quad \tilde{Z}_N \to^{a.s.}_{N\to\infty} \tilde{Z},$$

and we also know by Lemma 4.1.11 that $E\tilde{Z}_N^2 = E[Z_N^2|D_N(\omega)] = O(1)$ for all $\omega \in \Omega_0' \subset \Omega_0$ of probability $\Pr(\Omega_0') > 1 - \tau$ as close to one as desired. But this implies that the $(\tilde{Z}_N : N \in \mathbb{N})$ are uniformly integrable real random variables so that almost sure convergence implies convergence of first moments ([47], Theorem 10.3.6), that is

$$E|Z_n|D_N(\omega) - Z| = E|\tilde{Z}_N - \tilde{Z}| \to_{N\to\infty} 0$$

for all $\omega \in \Omega_0'$. In particular then, using also Fubini's theorem,

$$\sqrt{N}\big(\langle E^\Pi[\theta|D_N(\omega)], \psi\rangle_{L^2_\zeta} - \hat{\Psi}_N(\omega)\big) = E^\Pi\big[\sqrt{N}\big(\langle\theta, \psi\rangle - \hat{\Psi}_N\big)|D_N(\omega)\big] \to EZ = 0 \tag{4.19}$$

for $\omega \in \Omega_0'$. But if the last limit holds for all $\omega \in \Omega_0'$ with probability $\Pr(\Omega_0') > 1-\tau$ we have a contradiction to (4.18) (as then $\Pr(\Omega) \geq \Pr(\Omega') + \Pr(\Omega_0') > 1-\tau+\tau = 1$), completing the proof of (4.17) and thus of the theorem.

4.1.3 Applications to Uncertainty Quantification (UQ)

Let us illustrate a statistical application of the preceding theorems to Bayesian uncertainty quantification for functionals $\langle\theta, \psi\rangle_{L^2_\zeta(\mathcal{Z})}$. Consider a level $1-\xi$ Bayesian credible interval

$$C_N = \big\{v \in \mathbb{R} : |v - \langle\bar{\theta}, \psi\rangle_{L^2_\zeta(\mathcal{Z})}| \leq R_N/\sqrt{N}\big\}, \quad \Pi(C_N|D_N) = 1 - \xi, \tag{4.20}$$

where $\bar{\theta} = E^\Pi[\theta|D_N]$ is the posterior mean given the data $D_N = (Y_i, X_i)_{i=1}^N$ and $0 < \xi < 1$ a given coverage level. Construction of the interval C_N requires MCMC computation of that mean and of the quantiles R_N of the posterior distribution, but not of the asymptotic variance appearing in Theorem 4.1.10.

Now let $Q(t) = \Pr(|Z| \leq t), t \in \mathbb{R}$, where Z is the limiting normal distribution occurring in Theorem 4.1.10. Assume that Z is not degenerate (i.e., of non-zero variance) so that Q is continuous and strictly increasing. Then by Theorem 4.1.10 and the last part of Ex. 4.3.3 we see that

$$Q(R_N) = Q(R_N) - \Pi(C_N|D_N) + (1 - \xi) \to^{P^N_{\theta_0}} 1 - \xi$$

as $N \to \infty$ and since Q is invertible with continuous inverse, the continuous mapping theorem implies

$$R_N \to^{P_{\theta_0}^N} Q^{-1}(1-\xi).$$

By Slutsky's lemma and again Theorem 4.1.10, we then have as $N \to \infty$,

$$\frac{Q^{-1}(1-\xi)}{R_N}\sqrt{N}\langle \bar{\theta}_N - \theta_0, \psi\rangle_{L^2_\zeta(\mathcal{Z},W)} \to^d Z.$$

Next, let us compute the frequentist coverage probability of the credible set C_N for the true parameter θ_0. We have from the last limit and the continuous mapping theorem for $|\cdot|$ that

$$\begin{aligned}
P_{\theta_0}^N(\langle \theta_0, \psi\rangle_{L^2_\zeta} \in C_N) &= P_{\theta_0}^N(\sqrt{N}|\langle \theta_0 - \bar{\theta}_N, \psi\rangle_{L^2_\zeta}| \le R_N) \\
&= P_{\theta_0}^N\Big(\frac{Q^{-1}(1-\xi)}{R_N}\sqrt{N}|\langle \bar{\theta}_N - \theta_0, \psi\rangle_{L^2_\zeta}| \le Q^{-1}(1-\xi)\Big) \\
&\to Q(Q^{-1}(1-\xi)) = 1-\xi,
\end{aligned}$$

showing that C_N is indeed a precise asymptotic level $1-\xi$ confidence set of stochastic diameter $O(1/\sqrt{N})$. One can obtain 'simultaneous' versions of these results for appropriate collections of ψ's too if an approriate 'uniform' BvM theorem is at hand, see [61], Theorem 7.3.23.

4.2 Solving information equations in PDE models

The new condition encountered in this chapter concerns the solvability of the information equation (3.16). The corresponding Condition 4.1.2 does not only stipulate that a solution $\bar{\psi}$ exists, but also that it has sufficient regularity to belong to the space Θ. Whether this is possible or not is a priori not clear and depends on the particular inverse problem via the information operator (cf. Definition 3.1.2) induced by $\mathscr{G}$

For the PDE examples studied in these notes, our path reaches a perhaps surprising fork at this point: We will show in this section that for the Schrödinger model (2.5), the information operator is itself a composition of elliptic operators and can be inverted on the space $C_c^\infty(\mathcal{X})$ of smooth functions of compact support in $\mathcal{X}$ (or in fact on appropriate Sobolev spaces). As a consequence a Bernstein-von Mises theorem can be proved for $\langle \theta, \psi\rangle_{L^2}$ with $\theta \sim \Pi(\cdot|D_N)$ a posterior draw and for all $\psi \in C_c^\infty(\mathcal{X})$. These ideas extend to 'perturbed' differential operators $\mathscr{D} - f$ of the form mentioned after (1.4) – see the notes to this chapter for more discussion. In contrast, for Darcy's problem (2.8), we will show that the information equation can *not* be solved for typical elements of $C_c^\infty(\mathcal{X})$ (such as all *positive* smooth ψ).

In particular in view of Theorem 3.1.5 this implies that the Gaussian BvM-type approximations predicted by Le Cam theory for posterior functionals $\langle \theta, \psi \rangle$ do in fact *not* hold true for this PDE.

4.2.1 A Bernstein-von Mises theorem for the Schrödinger equation

We consider the Schrödinger forward map from (2.5) with potential $f = f_\theta = e^\theta, \theta \in \Theta$, and parameter space $C_c^\infty \subset \Theta \subset H^\xi(\mathcal{X}), \xi > d/2$. The information operator $\mathbb{I}_\theta^* \mathbb{I}_\theta$ is given by Theorem 3.3.1 and (3.54), (3.56). Now for $\psi \in C_c^\infty(\mathcal{X})$ and recalling the Schrödinger operator $\mathscr{L}_{f_\theta} = \frac{1}{2}\Delta - f_\theta$, we can define

$$\bar{\psi}_{\theta_0} = \frac{\mathscr{L}_{f_{\theta_0}} \mathscr{L}_{f_{\theta_0}} \left[\frac{\psi}{u_{f_{\theta_0}} f_{\theta_0}} \right]}{u_{f_{\theta_0}} f_{\theta_0}} \tag{4.21}$$

where we recall that $\mathscr{G}(\theta_0) = u_{f_{\theta_0}}$ is bounded away from zero on $\mathcal{X}$ (since $h \geq h_{min} > 0$ on $\mathcal{X}$ and using (6.29)). If we assume further that $\theta_0 \in C^\infty(\mathcal{X})$ then Proposition 6.1.5 and (4.21) imply that $\bar{\psi}_{\theta_0} \in C_c^\infty(\mathcal{X}) \in \Theta$.

Since $\mathscr{L}_{f_{\theta_0}}^{-1}$ is the inverse of $\mathscr{L}_{f_{\theta_0}}$ for Dirichlet boundary conditions (cf. (1.6) and (6.10)), we obtain from (3.54), (3.56) that

$$\begin{aligned} \mathbb{I}_{\theta_0}^* \mathbb{I}_{\theta_0} \bar{\psi}_{\theta_0} &= \mathbb{I}_{\theta_0}^* \left[\mathscr{L}_{f_{\theta_0}}^{-1} \mathscr{L}_{f_{\theta_0}} \mathscr{L}_{f_{\theta_0}} \left[\frac{\psi}{u_{f_{\theta_0}} f_{\theta_0}} \right] \right] \\ &= u_{f_{\theta_0}} f_{\theta_0} \mathscr{L}_{f_{\theta_0}}^{-1} \mathscr{L}_{f_{\theta_0}} \left[\frac{\psi}{u_{f_{\theta_0}} f_{\theta_0}} \right] = \psi \end{aligned} \tag{4.22}$$

for any $\psi \in C_c^\infty(\mathcal{X})$. That is, $\bar{\psi} = \bar{\psi}_{\theta_0}$ solves the information equation (3.16) for such ψ, and Condition 4.1.2 is verified. In particular, $(\mathbb{I}_{\theta_0}^* \mathbb{I}_{\theta_0})^{-1}$ is then a proper inverse on $C_c^\infty(\mathcal{X})$. Collecting these findings we have effectively proved the following Bernstein-von Mises theorem.

Theorem 4.2.1. *Consider a Gaussian prior Π_N from (2.22) with $\kappa = 0$ and base prior Π' as in Condition 2.2.1 where*

$$C_c^\infty \subset \Theta, \mathcal{H} \subset H^\alpha(\mathcal{X}), \mathcal{R} = H^\xi(\mathcal{X}), \xi > \alpha + d/2,$$

and with α satisfying

$$\frac{\alpha}{2\alpha + d} \frac{\alpha - d}{\alpha + 2 - d/2} > \frac{1}{3}. \tag{4.23}$$

Let $\theta \sim \Pi(\cdot|D_N)$ where $\Pi(\cdot|D_N)$ is the posterior measure (1.14) on Θ arising from observations $D_N = (Y_i, X_i)_{i=1}^N$ in model (1.11) with $\mathscr{G}(\theta)$ from (2.5). Denote the *posterior mean by* $\bar{\theta}_N = E^{\Pi}[\theta|D_N]$, *and let* $\psi \in C_c^\infty(\mathcal{X})$. *Assume* $\theta_0 \in C^\infty(\mathcal{X}) \cap \mathcal{H}$ *is such that* $\inf_{x\in\mathcal{X}} f_{\theta_0}(x) > f_{min}$. *Then we have*

$$\sqrt{N}\langle \theta - \bar{\theta}_N, \psi\rangle_{L^2_\lambda(\mathcal{X})}|D_N \to^d N(0, \sigma^2(f_{\theta_0}, \psi)) \quad \text{in } P^N_{\theta_0} - \text{probability},$$

as $N \to \infty$, *and moreover that*

$$\sqrt{N}\langle \bar{\theta}_N - \theta_0, \psi\rangle_{L^2_\lambda(\mathcal{X})} \to^d N(0, \sigma^2(f_{\theta_0}, \psi))$$

where the asymptotic variance is given by

$$\sigma^2(f_{\theta_0}, \psi) = \Big\| \mathscr{L}_{f_{\theta_0}}\Big[\frac{\psi}{u_{f_{\theta_0}} f_{\theta_0}}\Big]\Big\|^2_{L^2_\lambda(\mathcal{X})}. \tag{4.24}$$

Proof. We need to check conditions from Theorem 4.1.5. Condition 2.1.1 was checked for $\kappa = 0$ in Proposition 2.1.2, while Condition 2.1.4 was checked with $\eta = \xi/(2 + \xi)$ in Ex. 2.4.1. Condition 3.1.1 was verified in Theorem 3.3.1 while Condition 4.1.1 follows from Ex. 4.3.1. Condition 4.1.2 was checked just before the statement of the theorem. We can invoke Proposition 4.1.3 with $\beta = \xi, \kappa = 0$, and verify Condition 4.1.4 as follows: The $\|\cdot\|_\infty$-covering numbers of a β-Sobolev ball in dimension d are of the order

$$\log N(\Theta_{N,M,\infty}, \|\cdot\|_\infty, \epsilon) \lesssim \Big(\frac{1}{\epsilon}\Big)^{d/\beta}, \quad \beta > 0,$$

see (4.184) in [61] for the case when the functions are defined over $[0,1]^d$, and this bound applies to our setting by a standard extension argument (and regarding $\mathcal{X}$ as a subset of $[0,1]^d$ without loss of generality). Also for $\theta \in \Theta_{N,M,\infty}$ we can take $\rho_{\theta_0}(\theta - \theta_0 + (t/\sqrt{N})\bar{\psi}_{\theta_0}) \lesssim \bar{\delta}_N^2 \equiv \sigma_N$. We first note that the quantity in (4.7) is bounded by

$$\sqrt{N}\bar{\delta}_N^2 \int_0^1 (\bar{\delta}_N^2 \epsilon)^{-d/(2\beta)} d\epsilon \lesssim \sqrt{N}\bar{\delta}_N^{2-\frac{d}{\beta}} \tag{4.25}$$

since $\beta > d/2$. We will eventually show that the last bound converges to zero as $N \to \infty$, which also implies $N\sigma_N^2 \lesssim N\bar{\delta}_N^4 \to 0$. The middle term in the maximum in (4.8) can similarly be bounded by

$$\sqrt{N}\mathscr{J}_N(\sigma_N, 1) = \sqrt{N}\int_0^{\sigma_N} \epsilon^{-d/(2\beta)} d\epsilon \lesssim \sqrt{N}\bar{\delta}_N^{2-\frac{d}{\beta}},$$

and hence is of the same order as the one in (4.25). For the third member in the maximum (4.8) we have, by a similar calculation,

$$\frac{\bar{\delta}_N\sqrt{\log N}}{\sigma_N^2}\mathscr{J}_N^2(\sigma_N, 1) \lesssim \sqrt{\log N}\bar{\delta}_N^{1-\frac{2d}{\beta}}. \tag{4.26}$$

We can conclude from what precedes that it suffices to show that

$$\max\left(\sqrt{N}\bar{\delta}_N^{2-\frac{d}{\beta}}, N\bar{\delta}_N^3, \sqrt{\log N}\bar{\delta}_N^{1-\frac{2d}{\beta}}\right) \to 0 \tag{4.27}$$

as $N \to \infty$. This requires $\beta > 2d$ and then simplifies to the basic requirement $N\bar{\delta}_N^3 \to 0$. Using the convergence rate $\bar{\delta}_N$ provided by Proposition 4.1.3 and (4.23), Condition 4.1.4 is verified. The final result now follows from Theorems 4.1.5 and 4.1.10 and since, arguing as in the first step in (4.22),

$$\tilde{\psi}_{\theta_0} = \mathbb{I}_{\theta_0}\bar{\psi}_{\theta_0} = \mathscr{L}_{f_{\theta_0}}\Big[\frac{\psi}{u_{f_{\theta_0}}e^{\theta_0}}\Big]. \tag{4.28}$$

□

We note that the asymptotic variance in (4.24) is just the squared L^2_λ-norm of $\tilde{\psi}_{\theta_0}$ from (4.28) and hence precisely equals the lower bound from Theorem 3.1.4.

The previous result can in principle be made *uniform* in collections $\mathscr{W}$ of ψ's that are bounded in suitable Sobolev norms. If we wish to obtain such uniformity to hold for weak convergence (in $P^N_{\theta_0}$-probability) of the posterior measure towards the limiting centred Gaussian process $(\mathbb{X}(\psi) : \psi \in \mathscr{W})$ with covariance

$$E\mathbb{X}(\psi)\mathbb{X}(\psi') = \langle \mathscr{L}_{f_{\theta_0}}[\psi/(u_{f_{\theta_0}}f_{\theta_0})], \mathscr{L}_{f_{\theta_0}}[\psi'/(u_{f_{\theta_0}}f_{\theta_0})]\rangle_{L^2_\lambda},\ \psi, \psi' \in \mathscr{W}, \tag{4.29}$$

from (4.24), then this forces $\|\psi\|_{H^\beta} \le C$ for some $C > 0$ and $\beta > 2 + d/2$. The reason is that otherwise the limiting process would not be a tight random variable, so that by Prohorov's theorem, laws cannot converge weakly towards it – see Proposition 6 in [94] for a proof of this fact, and the notes to this section for more on the subject of 'uniform in ψ' Bernstein – von Mises theorems.

4.2.2 Impossibility of the BvM-phenomenon for Darcy's problem

We now show that for the forward map arising from the divergence form equation (2.8), even when considering the most regular parameter $\theta \equiv$ *const* corresponding to the standard Laplacian on a $2d$-disk, the information equation (3.16) can in fact *not be solved* for generic classes of smooth ψ's. The problem was already foreshadowed in the discussion before Proposition 2.1.5, only that now the Laplacian term that helped crucially in (2.15) has disappeared – the vector field ∇u_θ then entirely determines the solvability of the relevant information equation, and the counter-examples that follow exhibit that obstructions can arise from this in natural settings.

We will prove more than non-solvability of the information equation – we show that even the range condition $\psi \in R(\mathbb{I}^*_{\theta_0})$ is typically violated and that hence in view of Theorem 3.1.5, $\sqrt{N}$-consistent estimation of parameters $\langle \theta, \psi \rangle$ is not possible for a large class of smooth ψ's. In particular, *a fortiori*, no semi-parametric Bernstein - von Mises theorem can hold true, as this would entail the existence of an efficient estimator (whose asymptotic distribution coincides with the centring $\hat{\Psi}_N$ from (4.9)).

The range of $\mathbb{I}^*_\theta$ and transport PDEs

We consider the forward map $\mathscr{G}$ from (2.8) arising from solutions $\mathscr{G}(\theta) = u_\theta$ of the PDE (1.2) with divergence form operator (1.3) for $f = f_\theta = f_{min} + e^\theta, \theta \in \Theta = H^\beta_c(\mathcal{X})$. From Proposition 3.3.3 the range of the adjoint operator is given by

$$R(\mathbb{I}^*_\theta) = \left\{\psi = e^\theta \nabla u_\theta \cdot \nabla \mathscr{L}^{-1}_{f_\theta}[g], \text{ for some } g \in L^2_\lambda(\mathcal{X})\right\}, \tag{4.30}$$

The operator $\mathscr{L}^{-1}_{f_\theta}$ maps $L^2_\lambda(\mathcal{X})$ into $H^2_0(\mathcal{X}) = \{y \in H^2 : y_{|\partial\mathcal{X}} = 0\}$ and hence if ψ lies in the range of $\mathbb{I}^*_\theta$ then the equation

$$\begin{aligned} \nabla u_\theta \cdot \nabla y = e^{-\theta}\psi \quad &\text{on } \mathcal{X} \\ y = 0 \text{ on } &\partial\mathcal{X} \end{aligned} \tag{4.31}$$

necessarily has a solution $y = y_\psi \in H^2_0$. The existence of solutions to the transport PDE (4.31) depends in a possibly intricate way on the compatibility of ψ with geometric properties of the vector field ∇u_θ, which in turn is determined by the geometry of the forward map $\mathscr{G}$ (via f, g, θ) in the base PDE (1.2).

We now illustrate the obstructions that can arise from the above transport problem in the representative model case where $\mathscr{L}_{f_\theta}$ is the standard Laplacian Δ, i.e., $\theta = 0, f_{min} = 0$, and with source $g = 2$, boundary values $h = (|\cdot|^2_{\mathbb{R}^2} - 1)/2$. As underlying domain we choose either

i) $\mathcal{X}$ equal to a unit disk in $\mathbb{R}^2$ *separated away* from the origin $(0, 0)$, or
ii) $\mathcal{X}$ equal to the unit disk in $\mathbb{R}^2$ centred at $(0, 0)$.

We will refer to them as Example i) and ii), respectively, and note that the injectivity of $\mathbb{I}_\theta$ has already been studied in Ex. 3.4.5 and Proposition 3.3.4, respectively. By a simple perturbation argument the findings below extend to a neighbourhood of $\theta = 0$, see Ex. 4.3.4.

Example i)

In i), the solution to (1.2) with $f_\theta = 1$ equals $u_\theta = h$ and hence $\nabla u_\theta = x$ does not vanish on $\mathcal{X}$ for $\theta = 0$. The last property extends to any θ in a H^β-neighbourhood of the standard Laplacian $\theta = 0$ by Ex. 4.3.4.

The integral curves $\gamma(t)$ in $\mathcal{X}$ associated to the smooth vector field $\nabla u_\theta \neq 0$ are given near $x \in \mathcal{X}$ as the unique solutions (e.g., [122], p.9) of the vector ODE

$$\frac{d\gamma}{dt} = \nabla u_\theta(\gamma), \quad \gamma(0) = x. \tag{4.32}$$

Since ∇u_θ does not vanish we obtain through each $x \in \mathcal{X}$ a unique curve $(\gamma(t) : 0 \le t \le T_\gamma)$ originating and terminating at the boundary $\partial\mathcal{X}$, with finite 'travel time' $T_\gamma \le T(\mathcal{X}, c_\nabla) < \infty$. Along this curve the chain rule implies

$$\frac{d}{dt} y(\gamma(t)) = \frac{d\gamma(t)}{dt} \cdot \nabla y(\gamma(t)) = (\nabla u_\theta \cdot \nabla y)(\gamma(t)), \quad 0 < t < T_\gamma$$

and the PDE (4.31) reduces to the ODE

$$\frac{dy}{dt} = \psi. \tag{4.33}$$

Now suppose $\psi \in R(\mathbb{I}_\theta^*)$ then a solution $y \in H_0^2$ to (4.31) satisfying $y_{|\partial\mathcal{X}} = 0$ must exist. Such y then also solves the ODE (4.33) along each curve γ, with initial and terminal values $y(0) = y(T_\gamma) = 0$. By the fundamental theorem of calculus (and uniqueness of solutions) this forces

$$\int_0^{T_\gamma} \psi(\gamma(t))dt = 0 \tag{4.34}$$

to vanish. In other words, ψ permits a solution y to (4.31) only if ψ integrates to zero along each integral curve (orbit) induced by the vector field ∇u_θ. Now consider any smooth (non-zero) *nonnegative* $\psi \in C_c^\infty(\mathcal{X}) \subset \Theta$, and take $x \in \mathcal{X}$ such that $\psi \ge c > 0$ near x. For γ the integral curve passing through x we then cannot have (4.34) as the integrand never takes negative values while it is positive and continuous near x. Conclude by way of contradiction that $\psi \notin R(\mathbb{I}_\theta^*)$. Applying Theorem 3.1.5 with tangent space $H = C_c^\infty$ (and with its hypotheses verified in Theorem 3.3.2 and Ex. 4.3.1), we have proved.

Theorem 4.2.2. *Consider estimation of the functional* $\Psi(\theta) = \langle \theta, \psi \rangle_{L^2(\mathcal{X})}$ *from data* $(Y_i, X_i)_{i=1}^N$ *drawn i.i.d. from* P_θ^N *in the model (1.11) with forward map* $\mathscr{G}$ *from (2.8) with* $f_{min} = 1$*, where* g, h *in (1.2) are chosen as in Proposition 3.3.4, with the domain* $\mathcal{X}$ *separated away from the origin, and with tangent space* $H = H_c^\beta(\mathcal{X})$*. Suppose* $0 \neq \psi \in C_0^\infty(\mathcal{X})$ *satisfies* $\psi \ge 0$ *on* $\mathcal{X}$*. Then for* $\theta = 0$ *the efficient information for estimating* $\Psi(\theta)$ *satisfies*

$$\inf_{h \in H, \langle h, \psi \rangle_{L^2} \neq 0} \frac{\|\mathbb{I}_0 h\|_{L^2_\lambda}^2}{\langle \psi, h \rangle_{L^2_\lambda}^2} = 0. \tag{4.35}$$

In particular,

$$\liminf_{N\to\infty} \inf_{\bar\psi_N:(\mathbb{R}\times\mathcal{X})^N\to\mathbb{R}} \sup_{\theta'\in\Theta,\|\theta'\|_{L^2_\zeta}\le 1/\sqrt{N}} NE^N_{\theta'}(\bar\psi_N-\Psi(\theta'))^2=\infty. \tag{4.36}$$

The proof extends straightforwardly to a (fixed) neighborhood of the Laplacian $\theta=0$ as the key property used in the previous proof (that ∇u_θ is a non-vanishing vector field) extends by perturbation, see Ex. 4.3.4.

Let us notice that one can further show that (4.34) is also a *sufficient* condition for ψ to lie in the range of $\mathbb{I}_\theta^*$ (provided ψ is smooth and with compact support in $\mathcal{X}$). As this condition strongly depends on θ via the vector field ∇u_θ, it seems difficult to describe any choices of ψ that lie in $\cap_{\theta\in\Theta}R(\mathbb{I}_\theta^*)$.

The problem in this example is that while $\mathbb{I}_\theta$ is injective on $H_0^1\supset\Theta$, it is not injective any longer on the L^2-closure of Θ, which is the relevant issue for efficient estimation of functionals $\langle\theta,\psi\rangle$. The next counter-example is more subtle in this regard, as $\mathbb{I}_\theta$ is injective on all of L^2 in this case (see Proposition 3.3.4).

Example ii)

We now consider the same setting as in Theorem 4.2.2 but with $\mathcal{X}$ the unit disk centred at the origin, so that the vector field $\nabla u_\theta=\nabla u_0=x$ has a critical point at the origin when $\theta=0$ (again this property of an isolated zero extends by perturbation to a neighbourhood of θ, see Ex. 4.3.4).

We showed in Proposition 3.3.4 that for this example, the operator $\mathbb{I}_\theta$ is injective on all of $L^2(\mathcal{X})$, and hence any $\psi\in L^2(\mathcal{X})$ lies in the *closure* of the range of $\mathbb{I}_\theta^*$ (recalling the Hilbert space identity $\overline{R(\mathbb{I}_\theta^*)}=\ker(\mathbb{I}_\theta)^\perp$). Nevertheless, there are many relevant ψ's that are not contained in $R(\mathbb{I}_\theta^*)$ (before taking the closure). In the present example the gradient of u_θ vanishes and the integral curves γ associated to $\nabla u_\theta=(x_1,x_2)$ emanate along straight lines from $(0,0)$ towards boundary points $(z_1,z_2)\in\partial\mathcal{X}$, where necessarily $y((z_1,z_2))=0$ if y is to solve (4.31). If we parameterise these lines as

$$\{(z_1e^t,z_2e^t):-\infty<t\le 0\},$$

then as after (4.33) we see that if a solution $y\in H_0^2$ to (4.31) exists then ψ must necessarily satisfy

$$\int_{-\infty}^0 \psi(z_1e^t,z_2e^t)dt = 0-y(0)=const.\ \ \forall(z_1,z_2)\in\partial\mathcal{X}. \tag{4.37}$$

This again cannot happen, for example, for any non-negative non-zero $\psi\in H$ that vanishes along a given curve γ (for instance if it is zero in any given quadrant of $\mathcal{X}$), as this forces $const=0$. Theorem 3.1.5 again as in the preceding theorem yields the following:

Theorem 4.2.3. *Consider the setting of Theorem 4.2.2 but where now $\mathcal{X}$ is the unit disk centred at* $(0,0)$*, and where* $0 \leq \psi \in C_0^\infty(\mathcal{X}), \psi \neq 0$*, vanishes along some straight ray from* $(0,0)$ *to the boundary* $\partial\mathcal{X}$*. Then (4.35) and (4.36) hold at* $\theta = 0$*.*

Arguing as after Proposition 3.3.4, the result can be extended to any θ near the standard Laplacian by an application of the Morse lemma, see Ex. 4.3.4.

The last example reveals that it is the lack of a closed range of $\mathbb{I}_{\theta_0}$ that creates the obstruction for efficient inference here – in particular the non-existence of $\sqrt{N}$-consistent estimators for $\langle \theta, \psi \rangle$ is *not* a consequence of the lack of injectivity of $\mathbb{I}_\theta$. In the asymptotic setting of Le Cam theory we thus encounter fundamental limitations unless the tangent space H is strongly constrained (to a closed proper subspace of L^2_λ such as E_D from (5.3) for D fixed, cf. also the proof of Theorem 5.3.2 below). In the next chapter we will provide some partial 'non-asymptotic' remedies for this issue.

4.3 Notes

4.3.1 Exercises

Exercise 4.3.1. Show that Condition 4.1.1 holds for the forward map (2.5) arising with the Schrödinger equation. [Hint: Argue as in Proposition 2.1.2 and Theorem 3.3.1 but replace the L^2-continuity estimates for $\mathscr{L}_f^{-1}$ by L^∞-Lipschitz estimates from (6.31).] Show further that the inequality (3.6) is satisfied for the forward map (2.8) with $d \leq 3$ and any $h \in C^1$. [Hint: Adapt the estimate (2.11) via the Sobolev imbedding $H^2 \subset L^\infty$ and the elliptic regularity estimate (6.35).]

Exercise 4.3.2. Prove (4.11). [See p.142 in [127].]

Exercise 4.3.3. For μ_N, μ random probability measures on $\mathbb{R}$ defined on some probability space, suppose $\int_\mathbb{R} e^{tx} d\mu_N(x) \to \int_\mathbb{R} e^{tx} d\mu(x)$ in probability for all $t \in \mathbb{R}$. Show that $d_{weak}(\mu_N, \mu) \to 0$ in probability, and that the last limit also implies convergence to zero of the Kolmogorov distance $\sup_{x \in \mathbb{R}} |\mu_N((-\infty, x]) - \mu((-\infty, x])| \to 0$. [Hint: adapt the standard proof for non-random measures along almost surely convergent subsequences; also see the appendix of [30].]

Exercise 4.3.4. Suppose $\|\theta - 1\|_{H^\beta} < \eta$ for some $\beta > 1 + d/2$. Show that $\|u_\theta - u_1\|_\infty \to 0$ as $\eta \to 0$. [Hint: By the usual perturbation argument and the Sobolev imbedding,

$$\begin{aligned} \|u_\theta - u_1\|_{C^1} &\lesssim \|V_1[\nabla \cdot [(\theta - 1)\nabla u_\theta]]\|_{H^b} \lesssim \|(\theta - 1)\nabla u_\theta\|_{H^{b-1}} \\ &\lesssim \|\theta - 1\|_{H^{b-1}} \|u_\theta\|_{C^b} \leq \|\theta - 1\|_{H^\beta} \|u_\theta\|_{H^\beta} < C\eta.] \end{aligned} \tag{4.38}$$

Show further that if ∇u_1 has exactly one isolated critical point in $\mathcal{X}$, then so has u_θ for such θ and η small enough.

4.3.2 Remarks and comments

The results from this chapter on Bernstein-von Mises theorems are taken from [92] which build on earlier work in [94] as well as [90] and more generally with ideas for 'direct' models going back to [28–30] – see also [97, 99, 112]. The essential observation in [28, 29] to avoid the negative results due to [54, 72] was that in infinite-dimensions, Bernstein-von Mises theorems only hold in a weak enough topology, specifically weak enough such that the limiting Gaussian process is well-defined, cf. the discussion around (4.29). For inverse problems such 'functional' (uniform in ψ) results in negative order Sobolev-type spaces are obtained in [94] and [97, 99] based on earlier work in direct models [29] (see also [27, 31, 32] for work in settings not related to PDEs). The 'smooth test function' approach here with $\psi \in C_c^\infty$ allows for reasonably short proofs if smooth solutions to the information equation (3.16) exist (for smooth ψ), as is the case in Theorem 4.2.1.

The counter-examples for solvability of the information equation in Darcy's problem are due to [96]. They show that whether or not BvM theorems hold true depends in a delicate way on the PDE driving the inverse problem even in settings where ground truth and test function θ_0, ψ are both smooth (C^∞).

The invertibility of the information operator for the Schrödinger equation was shown in [94] and this extends to other examples where a base differential operator $\mathscr{D}$ is perturbed by some zero-order term ('potential'). A more involved example for which the theory of this chapter was shown to work in [92] are again the non-Abelian X-ray transforms from (1.1): For these $\mathbb{I}_\theta^*\mathbb{I}_\theta$ also turns out to be elliptic [106] at least in the interior of $\mathcal{X}$. Dealing with boundary issues at $\partial\mathcal{X}$ is a non-trivial task but the invertibilty of the information operator (with appropriate adjoint) was proved in [89, 90, 92].

Generally, when the information operator is elliptic then $\mathbb{I}_\theta^*$ can be expected to have closed range which allows the Le Cam machinery underlying the proof of Theorem 4.1.5 to proceed, whereas in other settings (such as Darcy's problem), the lack of closedness of the information operator appears to be at the heart of the issues pointed at here. These problems do not surface in the classical theory with fixed *finite-dimensional* tangent spaces H, as the range of a linear operator is then always closed and it suffices to prove injectivity of $\mathbb{I}_\theta$. In PDE inverse problems such a finite-dimensional setting can mask many subtleties encountered in high- and infinite-dimensions, see for instance [22] where it is shown that in the context of the Calderón problem from Section 1.1.1, when restricting to a fixed finite-dimensional space of piece-wise constant conductivities, a Bernstein-von Mises approximation does hold true at rate $1/\sqrt{N}$ (even though the non-parametric convergence rates for Hölder smooth conductivities are at best logarithmic [1]).

Chapter 5

Posteriors are probably log-concave

We now describe a distinct non-asymptotic perspective on the phenomenon that for large sample sizes, posterior measures arising from Gaussian process priors may resemble 'bell-shaped' or 'normal-type' distributions even in settings where $\mathscr{G}$ is non-linear and Θ is high-dimensional. We also investigate consequences of rigorous such results on the possibility of efficient computation of posterior distributions by MCMC methods.

We wish to remove some of the limitations of the Bernstein - von Mises type results from the previous Chapter. Theorems 4.2.2 and 4.2.3 show that in natural infinite-dimensional models Θ and representative PDE settings such as with Darcy's problem, smooth linear statistics of posterior measures may well *not* have an asymptotically Gaussian distribution. And even in settings where the posterior has an asymptotically normal distribution such as in the Schrödinger model considered in Theorem 4.2.1, the limiting Gaussian process obtained along the 'convergence determining class' of $\psi \in C_c^\infty(\mathcal{X})$ has a covariance structure that prevents the posterior measure to converge to its Gaussian limit except for in norms that are much weaker than L^2-norms – see the paragraph containing (4.29) for more discussion. So while such approximations are very exact and precise results, they hold only in possibly too restrictive scenarios.

It pays off though to take a second look at the issue: In the 'locally asymptotically normal (LAN)' world of Le Cam theory and semi-parametric statistics, one requires injectivity of $\mathbb{I}_\theta$ on all of L^2_ζ (see (3.15)) and needs exact solutions to the related information equation (3.16). On the other hand, from a non-asymptotic point of view, as discussed in Section 3.2, injectivity of $\mathbb{I}_\theta$ on an appropriate high-dimensional approximation space $E_D \subset \Theta$ is sufficient to obtain (average) local curvature of the log-likelihood function ℓ_N and then also of the posterior distribution. In particular an inequality such as (3.30) is sufficient and no exact solution of the information equation is required. This can make an essential difference as we illustrate in this section with Darcy's problem (where this issue was already partly foreshadowed in Exercise 3.4.5 and the discussion preceding Proposition 3.3.4).

A classical way to take advantage of such local curvature of ℓ_N would be to Taylor expand ℓ_N around an appropriate centring and to use the resulting quadratic approximation to construct a Gaussian proxy measure for $\Pi(\cdot|D_N)$ – these are called 'Laplace approximations' and will be discussed in more detail in Remark 5.1.4. We will take a different approach to this problem however and approximate the posterior density from the larger (itself infinite-dimensional) class of 'unimodal' *log-concave* distributions. These also explain the predicted 'bell-shaped' nature of posteriors but allow for faster approximation rates (measured in the relevant Wasserstein distance on probability measures) than Laplace approximations. They are also naturally compatible with gradient based MCMC methods, as we will see. Indeed, if a sufficiently good estimate of the centring is available, one can devise a Langevin MCMC algorithm that samples from the log-concave proxy and the Wasserstein approximation Theorem 5.1.3 implies that such methods perform well also as samplers for the actual posterior measure. In particular computation of the posterior mean will be shown to be possible in 'polynomial run-time' in relevant parameters. This holds for 'warm start' MCMC methods – the question of initialisation is discussed in the notes to this section. We also provide a result that illustrates why the problem of computing target measures by 'local' MCMC methods that are initialised outside a region of log-concavity is generally 'exponentially hard' in high-dimensional settings without additional structural assumptions. The results in this chapter show how one can leverage PDE structure and local log-concavity to break such computational hardness barriers.

5.1 Wasserstein approximation of the posterior

The idea in this section is to exploit the 'local curvature' Theorem 3.2.3 and contraction rate theorems such as Theorem 2.3.1 to show that the posterior measure from (1.14) is very well approximated by a log-concave (unimodal) distribution on high-dimensional approximation spaces $E_D \simeq \mathbb{R}^D$ of Θ that is concentrated near θ_0 (or rather its projection $\theta_{0,D}$ onto E_D). This will be true on the 'frequentist' event (subset of $(V \times \mathcal{X})^N$)

$$\begin{aligned}\mathcal{E}_{conv} := &\Big\{ \inf_{\theta\in\mathcal{B}} \lambda_{min}\big[-\nabla^2\ell_N(\theta)\big] \geq c_0 N D^{-\kappa_0}/2\Big\} \qquad (5.1)\\ &\cap \Big\{ \sup_{\theta\in\mathcal{B}} \Big[|\ell_N(\theta)| + \|\nabla\ell_N(\theta)\|_{\mathbb{R}^D} + \|\nabla^2\ell_N(\theta)\|_{op}\Big] \leq N(5c_1 D^{\kappa_1}+1)\Big\},\end{aligned}$$

whose probability was bounded in Theorem 3.2.3, and where we recall the set

$$\mathcal{B} = \mathcal{B}_{\mathbf{r}} = \big\{\theta \in E_D : \|\theta - \theta_{0,D}\|_{E_D} < \mathbf{r}\big\} \qquad (5.2)$$

from (3.31) for some radius $\mathbf{r}$ to be specified as well as the finite-dimensional approximation space $E_D \subset \Theta \subset L^2(\mathcal{Z})$ from (3.24). In fact, for conciseness in the proofs that follow we will choose the eigenbasis of the Dirichlet Laplacian Δ on the bounded smooth domain $\mathcal{Z} \subset \mathbb{R}^d$; for eigen-pairs $(\lambda_j, e_j) \in (0,\infty) \times H_0^1(\mathcal{Z})$ of the operator $-L_{1,0}$ from (6.9), we take

$$E_D = \Big\{\vartheta = \sum_{j=1}^{D} t_j e_j, t_j = \langle \vartheta, e_j\rangle_{L^2_\zeta(\mathcal{Z})}\Big\} \subset \Theta := \tilde{H}^\alpha(\mathcal{Z}),\ D \in \mathbb{N}, \tag{5.3}$$

with the the generalised Sobolev scale $\tilde{H}^\alpha(\mathcal{Z}) \subset H_0^\alpha(\mathcal{Z})$ defined in (6.15) below. While the results in this section *do* depend quantitatively on the action of the information operator from Definition 3.1.2 on the discretisation space E_D, generalisations beyond the choice of the eigenfunctions of the Dirichlet Laplacian (e.g., to wavelet bases on domains, or singular value decompositions of other operators with non-standard boundary conditions) are within reach as long as Condition (3.34) can be checked. See the notes to this section for more discussion.

The finite-dimensional approximation of Θ can be modelled explicitly in a Bayesian way by considering a 'sieved' prior truncated at some large enough $D \in \mathbb{N}$. For $\delta_N = N^{-\alpha/(2\alpha+d)}$, consider a Gaussian probability measure $\Pi_N = \mathrm{Law}(\theta)$ on E_D arising as

$$\theta = \frac{1}{\sqrt{N}\delta_N} \sum_{j=1}^{D} g_j \lambda_j^{-\alpha/2} e_j, \quad D \le N\delta_N^2,\ \alpha > 1 + d/2, \tag{5.4}$$

where the g_j's are i.i.d. $N(0,1)$. The posterior distribution from data in (1.11) then arises as in (1.14) and the contraction rate of $\Pi_N(\cdot|D_N)$ for general forward maps was already studied in Ex. 2.4.3.

We will require the 'average' curvature and regularity of the log-likelihood function that underpins Theorem 3.2.3 from earlier to hold on a sufficiently large neighbourhood of the projection $\theta_{0,D}$ of θ_0 onto E_D

Condition 5.1.1. *For $\mathcal{R} = \Theta = \tilde{H}^\alpha(\mathcal{Z}) \subset H_0^\alpha(\mathcal{Z})$ from (6.15) with $\alpha > 1 + d/2$, let $\mathscr{G}$ satisfy Condition 2.1.1 with $\kappa = 0$ and Condition 2.1.4 for some $0 < \eta \le 1$. Suppose further that Conditions 3.2.1 and 3.2.2 hold for some $\kappa_0, \kappa_1, \kappa_2 > 0, c_0, c_1, c_2 > 0$, and $\mathcal{B}, E_D \subset \Theta$ as in (5.2), (5.3) respectively, and with a choice of radius $\mathbf{r}$ satisfying, for $\delta_N = N^{-\alpha/(2\alpha+d)}$, that*

$$\mathbf{r} \ge (\log N)\tilde{\delta}_N, \quad \textit{where} \quad \tilde{\delta}_N \equiv (\log N)\max\big(\delta_N^\eta, D^{\kappa_0+\kappa_2}\delta_N\big). \tag{5.5}$$

The radius $\mathbf{r}$ on which 'average' curvature of ℓ_N is required for the main Theorem 5.1.3 below is thus determined by the 'local' and 'global' ill-posedness of the map $\mathscr{G}$, measured by the parameters κ_0 and η, respectively. In fact under somewhat stronger hypotheses (and by more involved proofs than those that follow), the term $D^{\kappa_0+\kappa_2}$ can be replaced by the smaller term $D^{\kappa_0/2}$ in (5.5), see the notes for discussion.

Under the above conditions it was shown in Theorem 3.2.3 that the event $\mathcal{E}_{conv}$ from (5.1) has $P_{\theta_0}^N$-probability of order $1 - O(\exp\{-CND^{-2\kappa_0-4\kappa_2}\})$. In words, the negative log-likelihood $-\ell_N(\theta)$ is strongly convex on $\mathcal{B}$ with high probability as long as the model dimension does is not too large (so that $ND^{-2\kappa_0-4\kappa_2} \to \infty$). Note that while one can choose D to diverge with N, the results in this section are 'non-asymptotic' and hold for all D bounded by constants to be specified. In particular the notation D_N remains reserved for the data vector (1.13).

5.1.1 Construction of a log-concave surrogate posterior

Given a choice of $\mathbf{r}$ from Condition 5.1.1 and the projection $\theta_{0,D}$ of θ_0 onto E_D, let θ_{init} be any fixed vector in E_D such that

$$\|\theta_{\text{init}} - \theta_{0,D}\|_{E_D} \le \mathbf{r}/8. \tag{5.6}$$

For the purposes of log-concave approximation theorems one may take $\theta_{\text{init}} = \theta_{0,D}$ itself but since $\theta_{0,D}$ is unknown, for applications to bounds on mixing times of 'feasible' MCMC type algorithms, a separate 'data-driven' initialiser θ_{init} is typically required. See the notes to this section for more discussion.

We require two auxiliary functions, $g_{\mathbf{r}}$ (with strongly convex tails) and $\alpha_{\mathbf{r}}$ (cut-off function): For some smooth and symmetric (about 0) $\varphi : \mathbb{R} \to [0,\infty)$ supported in $[-1,1]$ and integrating to $\int_{\mathbb{R}} \varphi(x)dx = 1$, let us define mollifiers $\varphi_h(x) := h^{-1}\varphi(x/h), h > 0$. Then, we define $\tilde{\gamma}_{\mathbf{r}}, \gamma_{\mathbf{r}} : \mathbb{R} \to \mathbb{R}$ by

$$\begin{aligned} \tilde{\gamma}_{\mathbf{r}}(t) &:= \begin{cases} 0 & \text{if} \quad t < 5\mathbf{r}/8, \\ (t - 5\mathbf{r}/8)^2 & \text{if} \quad t \ge 5\mathbf{r}/8, \end{cases} \\ \gamma_{\mathbf{r}}(t) &:= \big[\varphi_{\mathbf{r}/8} * \tilde{\gamma}_{\mathbf{r}}\big](t), \end{aligned} \tag{5.7}$$

where $*$ denotes convolution. Further, let $\alpha : [0,\infty) \to [0,1]$ be smooth and satisfy $\alpha(t) = 1$ for $t \in [0, 3/4]$ and $\alpha(t) = 0$ for $t \in [7/8, \infty)$. We then define $g_{\mathbf{r}} : E_D \to [0,\infty)$ and $\alpha_{\mathbf{r}} : E_D \to [0,1]$ as

$$g_{\mathbf{r}}(\theta) := \gamma_{\mathbf{r}}(\|\theta - \theta_{\text{init}}\|_{E_D}), \quad \alpha_{\mathbf{r}}(\theta) = \alpha\big(\|\theta - \theta_{\text{init}}\|_{E_D}/\mathbf{r}\big). \tag{5.8}$$

Now for ℓ_N as in (1.15) and K to be specified, define the 'surrogate' likelihood function

$$\tilde{\ell}_N(\theta) = \alpha_{\mathbf{r}}(\theta)\ell_N(\theta) - Kg_{\mathbf{r}}(\theta), \ \theta \in E_D, \tag{5.9}$$

which 'convexifies' $-\ell_N$ in the 'tails' as Proposition 5.1.2 shows. The proof is not difficult but somewhat technical and relegated to Ex. 5.4.1.

Proposition 5.1.2. *Let $\tilde{\ell}_N$ from (5.9) be defined for some constant K satisfying*

$$K \geq CN(c_1 D^{\kappa_1} + 1) \cdot (1 + \mathbf{r}^{-2}) \tag{5.10}$$

with c_1, κ_1 from Condition 5.1.1 and $C > 1$ depending only on the choice of the function α in (5.8). We then have

$$\ell_N(\theta) = \tilde{\ell}_N(\theta) \quad \textit{for all } \theta \in E_D \textit{ s.t. } \|\theta - \theta_{0,D}\|_{E_D} \leq 3\mathbf{r}/8. \tag{5.11}$$

Moreover, on the event $\mathcal{E}_{conv}$ from (5.1) we have that $\tilde{\ell}_N \in C^2(\mathbb{R}^D)$ (where $\mathbb{R}^D \simeq E_D$) and

$$\inf_{\theta \in E_D} \lambda_{min}\big(-\nabla^2 \tilde{\ell}_N(\theta)\big) \geq c_0 N D^{-\kappa_0}/2, \tag{5.12}$$

as well as

$$\|\nabla \tilde{\ell}_N(\theta) - \nabla \tilde{\ell}_N(\bar{\theta})\|_{E_D} \leq 7K \|\theta - \bar{\theta}\|_{E_D}, \quad \theta, \bar{\theta} \in E_D. \tag{5.13}$$

The resulting renormalised probability measure $\tilde{\Pi}(\cdot|(Y_i, X_i)_{i=1}^N) = \tilde{\Pi}(\cdot|D_N)$ will be our 'surrogate' posterior distribution. In light of the preceding proposition, on the high probability event $\mathcal{E}_{conv}$ this measure has a log-concave probability density on E_D given by

$$\tilde{\pi}(\theta|(Y_i, X_i)_{i=1}^N) = \frac{e^{\tilde{\ell}_N(\theta)}\pi(\theta)}{\int_{E_D} e^{\tilde{\ell}_N(\theta)}\pi(\theta)}, \quad \theta \in E_D, \tag{5.14}$$

where π is the probability density of the prior from (5.4).

5.1.2 The log-concave approximation theorem

The Wasserstein distance W_2 between probability measures μ and ν on E_D is

$$W_2^2(\mu, \nu) = \inf_{\tau} \int_{E_D \times E_D} |\theta - \theta'|_{E_D}^2 d\tau(\theta, \theta'), \tag{5.15}$$

where the infimum is taken over all 'couplings' τ of μ and ν, that is, probability measures τ on $E_D \times E_D$ whose marginals are μ, ν, respectively.

The following theorem establishes a quantitative bound on the log-concavity of $\tilde{\Pi}(\cdot|D_N)$, that the gradient of its log-density coincides with the gradient log-posterior near $\theta_{0,D}$ (and hence induces the same Langevin dynamics there, as is relevant in Section 5.2.1 to follow), and that in particular the posterior measure $\Pi(\cdot|D_N)$ is well approximated in Wasserstein distance by $\tilde{\Pi}(\cdot|D_N)$, with high probability under the law $P_{\theta_0}^N$ of the data $D_N = (Y_i, X_i)_{i=1}^N$. The ground truth $\theta_0 \in \Theta$

is assumed to belong to the Sobolev scale $\tilde{H}^\alpha(\mathcal{Z}) \supset E_D$ from (6.15), and is further assumed to be well approximated by its L^2-projection $\theta_{0,D}$ onto E_D;

$$\|\mathscr{G}(\theta_0) - \mathscr{G}(\theta_{0,D})\|_{L^2_\lambda} \leq \delta_N/2. \tag{5.16}$$

The latter condition can be checked easily using the L^2-Lipschitz property of $\mathscr{G}$ from Condition 2.1.1 combined with $\theta_0 \in \tilde{H}^\alpha(\mathcal{Z})$, if either θ_0 is smooth and/or D is large enough.

Theorem 5.1.3 (Log-concave Wasserstein-approximation)**.** *Let the spaces E_D and prior $\Pi = \Pi_N, \alpha$ be as in (5.3) and (5.4), respectively. Let the posterior distribution $\Pi(\cdot|D^N)$ from (1.14) arise from data $D_N = (Y_i, X_i)_{i=1}^N \sim P_\theta^N$ from (1.11) with forward map $\mathscr{G}$ from (1.10) satisfying Condition 5.1.1 for some $\mathbf{r} \leq 1$. Assume further that $D \leq A_1 N^{d/(2\alpha+d)}$ for some $A_1 < \infty$ and that $\theta_0 \in \tilde{H}^\alpha(\mathcal{Z})$ satisfies (5.16). Let the surrogate posterior measure $\tilde{\Pi}(\cdot|D_N)$ be as in (5.14) for the given value of $\mathbf{r}$ and any $K \geq N(\log N)D^{\kappa_1}/\mathbf{r}^2$.*

We then have the following:

A) For all $\theta \in \{\theta \in E_D : \|\theta - \theta_{0,D}\|_{E_D} \leq 3\mathbf{r}/8\}$ we have $\tilde{\ell}_N(\theta) = \ell_N(\theta)$ and

$$\nabla \log \tilde{\pi}(\cdot|D_N)(\theta) = \nabla \log \pi(\cdot|D_N)(\theta).$$

B) On an event $\mathcal{E}_N \subset (V \times \mathcal{X})^N$ of $P_{\theta_0}^N$-probability at least $1 - O(e^{-CND^{-2\kappa_0-4\kappa_2}})$ for some $C > 0$, the probability measure $\tilde{\Pi}(\cdot|D_N)$ is globally log-concave on E_D, specifically (with c_0 from Condition 3.2.2)

$$\inf_{\theta \in E_D} \lambda_{min}(-\nabla^2 \log \tilde{\pi}(\theta|D_N)) \geq \frac{c_0}{2} N D^{-\kappa_0},$$

and $\nabla \tilde{\ell}_N$ is globally Lipschitz on E_D with Lipschitz constant $7K$.

C) On an event $\mathcal{E}_N \subset (V \times \mathcal{X})^N$ of $P_{\theta_0}^N$-probability at least $1 - O(e^{-bN\delta_N^2} - e^{-CND^{-2\kappa_0-4\kappa_2}})$ for some $b > 0, C > 0$, we have the Wasserstein approximation

$$W_2^2\big(\Pi(\cdot|D_N), \tilde{\Pi}(\cdot|D_N)\big) \leq e^{-N^{d/(2\alpha+d)}}.$$

The theorem and the constants implicit in the O-notation are 'non-asymptotic' in the sense that whenever the hypotheses hold for pairs $(D, N) \in \mathbb{N}^2$, then so do the conclusions, (which are informative only when $N \to \infty$).

Remark 5.1.4 (Laplace approximations)**.** Let us remark that the log-concave approximations provided by the previous theorem are qualitatively different from commonly used 'Laplace approximations' for numerical integration of Gibbs measures (1.14). In such approximations the posterior is replaced by the Gaussian measure centred at a maximiser $\bar\theta$ of ℓ_N with inverse covariance equal to the Hessian of ℓ_N at $\bar\theta$ – see [135] and also [14, 67, 88, 116, 118] for recent references. In other words one replaces $\Pi(\cdot|D_N)$ by a measure with probability density

$$n(\theta|D_N) \propto \exp\left\{\frac{1}{2}(\theta-\bar\theta_N)^T\nabla^2\ell_N(\bar\theta)(\theta-\bar\theta_N)\right\}\pi(\theta), \quad \theta\in E_D,$$

which on the high probability event $\mathcal{E}_{conv}$ from (5.1) and for $\bar\theta\in\mathcal{B}$ gives a well-defined Gaussian approximation. Unlike in our log-concave approximation, the Hessian of $\log n(\cdot|D_N)$ is constant in θ. The numerical validity of such Laplace approximations is related to the stochastic order of the higher order terms in the Taylor series of ℓ_N about $\bar\theta$ and cannot in general be expected to result in convergence rates that are faster than algebraic in $1/N$. The added flexibility of being able to approximate from the (infinite-dimensional) class of *log-concave measures* instead of just from normal distributions hence can lead to substantially faster approximation rates such as those in Theorem 5.1.3C). Moreover, by virtue of Part A), the gradient of the log posterior remains the same in the region $\mathcal{B}$ where the bulk of the mass of $\Pi(\cdot|D_N)$ concentrates – the key statistical features of the posterior measure are hence not affected by the approximation. In contrast Laplace approximations do modify the posterior also in the core region $\mathcal{B}$.

Proof of Theorem 5.1.3

As mentioned at the outset of this section, there are two main ideas underpinning the proof of this theorem: One is Theorem 3.2.3 which implies that ℓ_N and then also the posterior distribution are *locally* log-concave near $\theta_{0,D}$, with high probability. The second ingredient is that the posterior measure is statistically consistent for θ_0 under $P_{\theta_0}^N$ (see Section 2.3, and Ex. 2.4.3 for the particular prior employed) and hence – by 5.5) – charges most of its mass precisely to the neighbourhood of $\theta_{0,D}$ where $\Pi(\cdot|(Y_i,X_i)_{i=1}^N)$ is log-concave. In particular the surrogate posterior differs from the real posterior only in an area of insignificant posterior mass, hence providing an accurate approximation. Making the last statement quantitative in high dimensions is the main challenge in the proof.

Part A) follows directly from (5.11) and the definition of $\tilde\pi(\cdot|(Y_i,X_i)_{i=1}^N)$ in (5.14). We next note that on the event $\mathcal{E}_{conv}$ from (5.1) and for our choice of K we can invoke Proposition 5.1.2 to verify the required gradient Lipschitz property in Part B) as well as

$$\sup_{\theta,\vartheta\in\mathbb{R}^D, \|\vartheta\|_{\mathbb{R}^D}=1} \vartheta^T[\nabla^2\tilde\ell_N(\theta)]\vartheta \le -\frac{c_0}{2}ND^{-\kappa_0}. \tag{5.17}$$

Part B) of Theorem 5.1.3 now follows from Theorem 3.2.3 bounding the probability of the event $\mathcal{E}_{conv}$ (cf. also after Condition 5.1.1). The rest of the proof is concerned with part C), using Part A) and B) and w.l.o.g. for all N large enough.

Decomposition of the Wasserstein distance.

For this proof let us parameterise the Euclidean balls

$$\mathcal{B}(r) = \{\theta \in E_D : \|\theta - \theta_{0,D}\|_{E_D} \le r\} \text{ by their radius } r > 0.$$

By Theorem 6.15 in [133] with $x_0 = \theta_{0,D}$

$$W_2^2(\tilde{\Pi}(\cdot|D_N), \Pi(\cdot|D_N)) \le 2 \int_{E_D} \|\theta - \theta_{0,D}\|_{E_D}^2 d|\tilde{\Pi}(\cdot|D_N) - \Pi(\cdot|D_N)|(\theta).$$

This can be further bounded, for $m > 0$ to be chosen and $\tilde{\delta}_N$ from Condition 5.1.1, by

$$\begin{aligned} &\le 2m^2\tilde{\delta}_N^2 \int_{\mathcal{B}(m\tilde{\delta}_N)} d|\Pi(\cdot|D_N) - \tilde{\Pi}(\cdot|D_N)|(\theta) \\ &\quad + 2\int_{\|\theta-\theta_{0,D}\|_{E_D} > m\tilde{\delta}_N} \|\theta - \theta_{0,D}\|_{E_D}^2 d\tilde{\Pi}(\theta|D_N) \\ &\quad + 2\int_{\|\theta-\theta_{0,D}\|_{E_D} > m\tilde{\delta}_N} \|\theta - \theta_{0,D}\|_{E_D}^2 d\Pi(\theta|D_N) \equiv I + II + III, \end{aligned}$$

and we now bound I, II, III in separate steps.

Term II. We write the surrogate posterior density (5.14) as

$$\tilde{\pi}(\theta|D_N) = \frac{e^{\tilde{\ell}_N(\theta) - \tilde{\ell}_N(\theta_{0,D})}\pi(\theta)}{\int_{E_D} e^{\tilde{\ell}_N(\theta) - \tilde{\ell}_N(\theta_{0,D})}\pi(\theta)d\theta}, \quad \theta \in E_D.$$

Since $\delta_N = o(\tilde{\delta}_N)$ we have for $\mathbf{r}$ from Condition 5.1.1 and some $c = c(L) > 0$ small enough (with $L = L_{\mathscr{G}}(1)$ the local Lipschitz constant from (2.4)), the set inclusion

$$B_N := \{\|\theta - \theta_{0,D}\|_{E_D} \le c\delta_N\} \subset \big\{\|\theta - \theta_{0,D}\|_{E_D} \le 3\mathbf{r}/8\big\} \cap \big\{\|\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\|_{L^2_\lambda} \le \delta_N/2\big\},$$

using also (5.16). In particular $\ell_N(\theta) = \tilde{\ell}_N(\theta)$ on B_N (cf. (5.11)). Next note that a version of Lemma 1.3.3 with $\ell_N(\theta_0)$ replaced by $\ell_N(\theta_{0,D})$ is proved analogously, using (5.16) to control projection terms. Applying the so modified lemma with $\nu = \Pi(\cdot)/\Pi(B_N)$ as well as the small ball estimate given in step ii) in the proof of Theorem 2.2.2 (with $\kappa = 0$ and adapted to the sieved case as in Ex. 2.4.3), we deduce that for $\bar{c} = 2 + A$

$$\begin{aligned} \int_{E_D} e^{\tilde{\ell}_N(\theta) - \tilde{\ell}_N(\theta_{0,D})} d\Pi(\theta) &\ge \int_{B_N} e^{\tilde{\ell}_N(\theta) - \tilde{\ell}_N(\theta_{0,D})} d\Pi(\theta) \qquad (5.18) \\ &= \int_{B_N} e^{\ell_N(\theta) - \ell_N(\theta_{0,D})} d\nu(\theta) \times \Pi(B_N) \ge e^{-\bar{c}N\delta_N^2} \end{aligned}$$

on events $\mathcal{A}_N$ of $P^N_{\theta_0}$-probability approaching one (in fact, of the required order $1 - O(e^{-bN\delta_N^2})$, if we increase $\bar{c}$ appropriately, see Ex. 5.4.2).

Then on the preceding events the term II can be bounded, using a second order Taylor expansion of $\tilde{\ell}_N(\theta)$ around $\theta_{0,D}$ with $\bar{\theta}$ lying on the line segment connecting θ and $\theta_{0,D}$, as

$$\int_{\|\theta-\theta_{0,D}\|_{E_D}>m\tilde{\delta}_N} \|\theta-\theta_{0,D}\|^2_{E_D} \tilde{\pi}(\theta|D_N)d\theta \leq$$
$$e^{\bar{c}N\delta_N^2} \int_{\|\theta-\theta_{0,D}\|_{E_D}>m\tilde{\delta}_N} \|\theta-\theta_{0,D}\|^2_{E_D} e^{\tilde{\ell}_N(\theta)-\tilde{\ell}_N(\theta_{0,D})}\pi(\theta)d\theta =$$
$$e^{\bar{c}N\delta_N^2} \int_{\|\theta-\theta_{0,D}\|_{E_D}>m\tilde{\delta}_N} \|\theta-\theta_{0,D}\|^2_{E_D} e^{\nabla\tilde{\ell}_N(\theta_{0,D})^T(\theta-\theta_{0,D})+(\theta-\theta_{0,D})^T\nabla^2\tilde{\ell}_N(\bar{\theta})(\theta-\theta_{0,D})/2}\pi(\theta)d\theta.$$

For the first term in the exponent in the last integral we have (since $\tilde{\ell}_N = \ell_N$ at $\theta_{0,D}$), using the Cauchy-Schwarz inequality,

$$\begin{aligned} &|\nabla\tilde{\ell}_N(\theta_{0,D})^T(\theta-\theta_{0,D})| \\ &\leq \big(\|\nabla\ell_N(\theta_{0,D}) - E_{\theta_0}\nabla\ell_N(\theta_{0,D})\|_{\mathbb{R}^D} + \|E_{\theta_0}\nabla\ell_N(\theta_{0,D})\|_{\mathbb{R}^D}\big)\|\theta-\theta_{0,D}\|_{E_D}. \end{aligned} \tag{5.19}$$

In view of (3.39) and Condition 3.2.1 we can apply Bernstein's inequality (6.48) to the first summand to deduce

$$\|\nabla\ell_N(\theta_{0,D}) - E_{\theta_0}\nabla\ell_N(\theta_{0,D})\|^2_{\mathbb{R}^D} \lesssim DN \tag{5.20}$$

with sufficiently high probability (see Ex. 5.4.2 for details). Likewise, E_{θ_0}-integrating the identity (3.39) and using Conditions 3.2.1, (5.16), the second summand is of order

$$\|E_{\theta_0}\nabla\ell_N(\theta_{0,D})\|_{\mathbb{R}^D} \lesssim ND^{\kappa_2}\|\mathscr{G}(\theta_{0,D}) - \mathscr{G}(\theta_0)\|_{L^2_\lambda} \lesssim ND^{\kappa_2}\delta_N.$$

Hence using the hypothesis on D, the r.h.s. in (5.19) is further upper bounded by

$$(\sqrt{DN} + ND^{\kappa_2}\delta_N)\|\theta-\theta_{0,D}\|_{E_D} \lesssim ND^{\kappa_2}\delta_N\|\theta-\theta_{0,D}\|_{E_D}.$$

Now by virtue of (5.17) the exponent in the integrand above (5.19) is bounded by

$$\Big(\bar{L}\frac{ND^{\kappa_2}\delta_N}{\|\theta-\theta_{0,D}\|_{\mathbb{R}^D}} - \frac{c_0}{2}ND^{-\kappa_0}\Big)\|\theta-\theta_{0,D}\|^2_{E_D} \tag{5.21}$$

for some $\bar{L} > 0$. Then for m large enough and by definition of $\tilde{\delta}_N$ from Condition 5.1.1 we can bound this further from above by

$$-\frac{c_0}{4}ND^{-\kappa_0}\|\theta-\theta_{0,D}\|^2_{E_D}$$

for θ in the relevant domain of integration, so that the last displayed term above (5.19) is bounded by

$$e^{\bar{c}N\delta_N^2}e^{-\frac{c_0m^2}{4}ND^{-\kappa_0}\tilde{\delta}_N^2}E^{\Pi}[\|\theta-\theta_{0,D}\|_{E_D}^2].$$

The second moments of the Gaussian measure Π_N are uniformly bounded, and $\|\theta_{0,D}\|_{E_D}$ is uniformly bounded by a constant depending only on $\|\theta_0\|_{L^2}$. From Condition 5.1.1 we conclude that the term II does not exceed $e^{-N\delta_N^2}/4$ with sufficiently high probability.

Term III: We first note that from Ex. 2.4.3 and (5.16) we can prove (2.33) and then (2.31) with $\theta_{0,D}$ replacing θ_0 there, that is, the posterior contracts about $\theta_{0,D}$ just as about θ_0. This observation and (5.5) then imply that for every $a>0$ we can find m large enough such that

$$\Pi(\|\theta-\theta_{0,D}\|_{E_D}>m\tilde{\delta}_N|D_N)\le e^{-aN\delta_N^2}$$

on events $\mathcal{S}'_N$ of sufficiently high $P_{\theta_0}^N$-probability. We can further restrict to $\mathcal{A}_N$ on which the last inequality in (5.18) holds (see also Ex. 5.4.2). Then using the Cauchy-Schwarz inequality, the identity $E_{\theta_0}^N e^{\ell_N(\theta)-\ell_N(\theta_0)}=1$ as in (1.39) and Markov's inequality, we have

$$\begin{aligned}
&P_{\theta_0}^N\Big(\mathcal{A}_N\cap\mathcal{S}'_N,\int_{\|\theta-\theta_{0,D}\|_{E_D}>m\tilde{\delta}_N}\|\theta-\theta_{0,D}\|_{E_D}^2d\Pi(\cdot|D_N)>e^{-N\delta_N^2}/8\Big)\le\\
&P_{\theta_0}^N\Big(\mathcal{A}_N\cap\mathcal{S}'_N,\Pi(\|\theta-\theta_{0,D}\|_{E_D}>m\tilde{\delta}_N|D_N)E^{\Pi}[\|\theta-\theta_{0,D}\|_{E_D}^4|D_N]>\frac{e^{-2N\delta_N^2}}{64}\Big)\\
&\le P_{\theta_0}^N\Big(\mathcal{S}'_N,e^{(2+\bar{c}-a)N\delta_N^2}\int_{E_D}\|\theta-\theta_{0,D}\|_{E_D}^4e^{\ell_N(\theta)-\ell_N(\theta_0)}d\Pi(\theta)>1/64\Big)\\
&\lesssim e^{(2+\bar{c}-a)N\delta_N^2}\int_{E_D}(1+\|\theta\|_{E_D}^4)d\Pi(\theta)=O(e^{-bN\delta_N^2})
\end{aligned}$$

whenever m and then a are large enough, since Π has uniformly bounded fourth moments and since $\|\theta_{0,D}\|_{E_D}$ is uniformly bounded by a constant depending only on $\|\theta_0\|_{L^2}$. This shows that term III is of the required order with sufficiently high probability.

Term I: We have from Condition 5.1.1 that for fixed $m>0$ and all N large enough

$$\mathcal{B}(m\tilde{\delta}_N)\subseteq\{\theta:\|\theta-\theta_{0,D}\|_{E_D}\le 3\mathbf{r}/8\}.$$

On the latter set, by (5.11), the probability measures $\tilde{\Pi}(\cdot|D_N)$ and $\Pi(\cdot|D_N)$ coincide up to a normalising factor, and thus we can represent their Lebesgue densities as

$$\tilde{\pi}(\theta|D_N)=p_N\pi(\theta|D_N),\quad\theta\in\mathcal{B}(m\tilde{\delta}_N),$$

for some $0 < p_N < \infty$. Moreover, by the preceding estimates for terms II and III (which hold just as well without the integrating factors $n(\theta) := \|\theta - \theta_{0,D}\|^2_{\mathbb{R}^D}$, as only $E^\Pi n(\theta) \leq const$ was used), we have

$$p_N \Pi(\mathcal{B}(m\tilde{\delta}_N)|D_N) = \tilde{\Pi}(\mathcal{B}(m\tilde{\delta}_N)|D_N) \geq 1 - e^{-N\delta_N^2}/8 \ \Rightarrow \ 1 - e^{-N\delta_N^2}/8 \leq p_N,$$

$$p_N^{-1} \tilde{\Pi}(\mathcal{B}(m\tilde{\delta}_N)|D_N) = \Pi(\mathcal{B}(m\tilde{\delta}_N)|D_N) \geq 1 - e^{-N\delta_N^2}/8 \ \Rightarrow \ 1 - e^{-N\delta_N^2}/8 \leq \frac{1}{p_N}$$

on events of sufficiently high $P_{\theta_0}^N$-probability. On these events necessarily

$$p_N \in \Big[1 - \frac{e^{-N\delta_N^2}}{8}, \big[1 - \frac{e^{-N\delta_N^2}}{8}\big]^{-1}\Big]$$

and so for N large enough

$$\begin{aligned}
&\int_{\mathcal{B}(m\tilde{\delta}_N)} d|\Pi(\cdot|D_N) - \tilde{\Pi}(\cdot|D_N)|(\theta) \\
&= |1 - p_N| \int_{\mathcal{B}(m\tilde{\delta}_N)} \pi(\theta|D_N) d\theta \leq |1 - p_N| \leq e^{-N\delta_N^2}/4,
\end{aligned}$$

which is obvious for $p_N \leq 1$ as then $|1 - p_n| = 1 - p_n$ and follows from the mean value theorem applied to $f(x) = (1-x)^{-1}$ near $x = 0$ also for $p_N > 1$. Collecting all the bounds we have shown that $I + II + III \leq e^{-N\delta_N^2}$ with the desired $P_{\theta_0}^N$-probability, completing the proof of part C) of Theorem 5.1.3.

5.2 Computational complexity of MCMC in high dimensions

5.2.1 Gradient methods for approximately log-concave posteriors

We now consider the problem of generating random samples from a high-dimensional posterior measure

$$\Pi(B|D_N) = \frac{\int_B e^{\ell_N(\theta)} d\Pi(\theta)}{\int_{E_D} e^{\ell_N(\theta)} d\Pi(\theta)}, \quad B \subseteq E_D \text{ measurable}, \tag{5.22}$$

arising from data (1.11) with log-likelihood function (1.15) and a general Gaussian $N(0, \Sigma)$ prior Π of density π on $\mathbb{R}^D \simeq E_D \subseteq \Theta$. We specifically have the prior (5.4) in mind where $\Sigma = diag(\lambda_j^\alpha : j \leq D)$ but the results in this subsection hold for general Gaussian priors with positive definite covariance matrix $\Sigma \in \mathbb{R}^{D\times D}$, and do not depend on the particular choice of E_D from (5.3) either.

We use the stochastic gradient method (1.21) obtained from discretisation of the D-dimensional Langevin diffusion (1.20) with drift vector field $\nabla(\tilde{\ell}_N + \log \pi)$ based on the *surrogate* likelihood function $\tilde{\ell}_N$ from (5.9), centred at θ_{init} from (5.6). More precisely, for *stepsize* $\gamma > 0$ and auxiliary variables $\xi_k \sim^{i.i.d.} N(0, I_{D\times D})$, define a Markov chain

$$\begin{aligned} \vartheta_0 &= \theta_{\text{init}}, \\ \vartheta_{k+1} &= \vartheta_k + \gamma\big[\nabla\tilde{\ell}_N(\vartheta_k) - \Sigma^{-1}\vartheta_k\big] + \sqrt{2\gamma}\xi_{k+1}, \quad k = 0, 1, \dots \end{aligned} \tag{5.23}$$

Probabilities and expectations with respect to the law of this Markov chain (random only through the ξ_k, conditional on the data D_N) will be denoted by $\mathbf{P}_{\theta_{\text{init}}}, \mathbf{E}_{\theta_{\text{init}}}$ respectively. The invariant measure of the underlying continuous time Langevin diffusion equals the *surrogate posterior distribution* $\tilde{\Pi}(\cdot|D_N)$ with density $\tilde{\pi}(\cdot|D_N)$ given in (5.14).

The idea of the above algorithm is that if we can initialise into the region $\mathcal{B}$ of local average log-concavity from (5.2), then we can force the gradient method back to $\mathcal{B}$ whenever it exits that region, by virtue of the dynamics of the 'surrogate' vector field $\nabla\tilde{\ell}_N$ (cf. Proposition 5.1.2). By construction, the dynamics of the above MCMC scheme coincides exactly with the standard Langevin method targeting $\Pi(\cdot|D_N)$ in the region $\mathcal{B}$ where the bulk of the posterior mass lies.

The mixing time of this Markov chain towards its (with high probability, log-concave) surrogate target measure measured in Wasserstein distance can be controlled by high-dimensional Markov chain theory (see Section 6.2.3), and Theorem 5.1.3 controls the approximation error induced by the surrogate approximation. Formally, the results that follow will apply the results from Section 6.2.3 on an intersection

$$\mathcal{E} := \mathcal{E}_{conv} \cap \mathcal{E}_{wass}(\rho), \tag{5.24}$$

of frequentist 'high probability' events (i.e., measurable subset of $(V \times \mathcal{X})^N$): first we restrict to the 'curvature' event $\mathcal{E}_{conv}$ from (5.1) controlled in Theorem 3.2.3 and we further assume that the Wasserstein distance W_2 between $\tilde{\Pi}(\cdot|D_N)$ and $\Pi(\cdot|D_N)$ can be bounded, specifically, for any $\rho > 0$, let us define the event

$$\mathcal{E}_{wass}(\rho) := \big\{W_2^2\big(\Pi(\,\cdot\,|D_N), \tilde{\Pi}(\,\cdot\,|D_N)\big) \le \rho/2\big\}. \tag{5.25}$$

See Theorem 5.1.3C) for sufficient conditions for this to be the case with 'small' ρ decaying exponentially in N.

Our first result consists of a global Wasserstein-approximation of $\Pi(\cdot|D_N)$ by the law $\mathcal{L}(\vartheta_k)$ on $\mathbb{R}^D$ of the k-th iterate ϑ_k arising from (5.23). Note that the event

$\mathcal{E}$ from Theorem 3.2.3 on which it holds implicitly stipulates choices of the constants $\kappa_i, i = 0, 1, 2$, describing analytical properties of the average log-likelihood function.

Theorem 5.2.1 (Wasserstein mixing time)**.** *Suppose that $D, N \in \mathbb{N}$ are such that $D \le \mathcal{R}_N$ with $\mathcal{R}_N = CND^{-2\kappa_0-4\kappa_2}$ from (3.36) and let K be as in (5.10). Further, if $\lambda_{min}(\cdot)$ and $\lambda_{max}(\cdot)$ define the smallest and largest eigenvalues of a symmetric matrix, respectively, define the constants*

$$m := \frac{Nc_0 D^{-\kappa_0}}{2} + \lambda_{min}(\Sigma^{-1}), \qquad \Lambda := 7K + \lambda_{max}(\Sigma^{-1}).$$

Then for any $0 < \gamma \le 1/\Lambda$ and any $\rho > 0$ the algorithm $(\vartheta_k : k \ge 0)$ from (5.23) initialised at θ_{init} satisfying (5.6) satisfies, on the event $\mathcal{E}$ from (5.24) and for all $k \ge 0$,

$$W_2^2\big(\mathcal{L}(\vartheta_k), \Pi(\cdot|D_N)\big) \le \rho + b(\gamma) + 4\Big(\tau(\Sigma, R) + \frac{D}{m}\Big)\Big(1 - \frac{\gamma m}{2}\Big)^k, \tag{5.26}$$

where, for some universal constants $c_1, c_2 > 0$, any $R \ge \|\theta_{0,D}\|_{E_D}$ and $\kappa(\Sigma) = \lambda_{max}(\Sigma)/\lambda_{min}(\Sigma)$,

$$b(\gamma) = c_1\Big[\frac{\gamma D\Lambda^2}{m^2} + \frac{\gamma^2 D\Lambda^4}{m^3}\Big], \quad \tau(\Sigma, R) = c_2\kappa(\Sigma)\Big[1 + \mathbf{r}^2 + R^2\Big]. \tag{5.27}$$

Proof. For any $\theta, \bar\theta \in E_D$, we have for the log-prior density that

$$\begin{aligned}\|\nabla \log \pi(\theta) - \nabla \log \pi(\bar\theta)\|_{E_D} = \|\Sigma^{-1}(\theta - \bar\theta)\|_{E_D} &\le \lambda_{max}(\Sigma^{-1})\|\theta - \bar\theta\|_{E_D},\\ \lambda_{min}(-\nabla^2 \log \pi(\theta)) &\ge \lambda_{min}(\Sigma^{-1}),\end{aligned}$$

and for the likelihood surrogate $\tilde\ell_N$, by Proposition 5.1.2 and on the event $\mathcal{E}$ from (5.24), that

$$\begin{aligned}\|\nabla\tilde\ell_N(\theta) - \nabla\tilde\ell_N(\bar\theta)\|_{E_D} &\le 7K\|\theta - \bar\theta\|_{E_D},\\ \lambda_{min}(-\nabla^2\tilde\ell_N(\theta)) &\ge Nc_0D^{-\kappa_0}/2.\end{aligned}$$

Combining the last two displays, and on the event $\mathcal{E}$, we can verify Assumption 6.2.5 below for $-\log d\tilde\Pi(\cdot|D_N)$ with constants

$$m = \frac{Nc_0 D^{-\kappa_0}}{2} + \lambda_{min}(\Sigma^{-1}), \qquad \Lambda = 7K + \lambda_{max}(\Sigma^{-1}).$$

We may thus apply Proposition 6.2.6 below to obtain,

$$\begin{aligned}W_2^2(\mathcal{L}(\vartheta_k), \Pi(\cdot|D_N)) &\le 2W_2^2(\Pi(\cdot|D_N), \tilde\Pi(\cdot|D_N)) + 2W_2^2(\mathcal{L}(\vartheta_k), \tilde\Pi(\cdot|D_N))\\ &\le \rho + b(\gamma) + 4(1 - m\gamma/2)^k\Big[\|\theta_{\mathrm{init}} - \theta_{max}\|_{E_D}^2 + \frac{D}{m}\Big],\end{aligned}$$

where θ_{max} denotes the unique maximiser of $\log d\tilde{\Pi}(\cdot|D_N)$ over $\mathbb{R}^D$ (which exists on the event $\mathcal{E}_{conv}$, by virtue of strong concavity).

We conclude by an estimate for $\|\theta_{\rm init} - \theta_{max}\|_{E_D}$. One may expect this norm to be small compared to $\mathbf{r}$, but at this point a crude estimate will suffice. From (5.8) we have for all $\|\theta - \theta_{\rm init}\|_{E_D} \geq 2\mathbf{r}$ that $g_{\mathbf{r}}(\theta) \geq \big(\|\theta - \theta_{\rm init}\|_{E_D} - \mathbf{r}\big)^2 \geq \frac{1}{4}\|\theta - \theta_{\rm init}\|_{E_D}^2$. Thus, for C from (5.10) and any $\theta \in E_D$ satisfying

$$\|\theta - \theta_{\rm init}\|_{E_D}^2 \geq \frac{20}{C} + 4\mathbf{r}^2,$$

and using also the upper bound for $|\ell_N(\theta)|$ in the definition of $\mathcal{E}_{conv}$, we obtain

$$\begin{aligned} -\tilde{\ell}_N(\theta) = Kg_{\mathbf{r}}(\theta) &\geq CN(c_1D^{\kappa_1}+1)(1+1/\mathbf{r}^2)\cdot \frac{\|\theta - \theta_{\rm init}\|_{E_D}^2}{4} \\ &\geq \frac{C}{4}N(c_1D^{\kappa_1}+1)\|\theta - \theta_{\rm init}\|_{E_D}^2 \\ &\geq 5N(c_1D^{\kappa_1}+1) \geq -\tilde{\ell}_N(\theta_{\rm init}). \end{aligned}$$

This implies that necessarily the unique maximiser $\theta_{\tilde{\ell}}$ of the (on $\mathcal{E}_{conv}$) strongly concave map $\tilde{\ell}_N$ over $\mathbb{R}^D$ satisfies $\|\theta_{\tilde{\ell}} - \theta_{\rm init}\|_{\mathbb{R}^D}^2 \leq 20/C + 4\mathbf{r}^2$. Moreover, in view of the definition of $\mathcal{B}$ we have that

$$\|\theta_{\rm init}\|_{E_D} \leq \|\theta_{\rm init} - \theta_{0,D}\|_{E_D} + \|\theta_{0,D}\|_{E_D} \leq \|\theta_{\rm init} - \theta_{0,D}\|_{E_D} + R \leq \mathbf{r} + R,$$

which also allows us to deduce

$$\|\theta_{\tilde{\ell}}\|_{E_D} \leq \|\theta_{\tilde{\ell}} - \theta_{\rm init}\|_{E_D} + \|\theta_{\rm init}\|_{E_D} \leq \sqrt{20/C} + 3\mathbf{r} + R.$$

We further have that $\theta_{max}^T\Sigma^{-1}\theta_{max} \leq \theta_{\tilde{\ell}}^T\Sigma^{-1}\theta_{\tilde{\ell}}$ (otherwise θ_{max} would not maximise $\log d\tilde{\Pi}(\cdot|D_N)$) and thus, for $\kappa(\Sigma)$ the condition number of Σ,

$$\|\theta_{max}\|_{E_D}^2 \leq \frac{1}{\lambda_{min}(\Sigma^{-1})}\theta_{max}^T\Sigma^{-1}\theta_{max} \leq \frac{1}{\lambda_{min}(\Sigma^{-1})}\theta_{\tilde{\ell}}^T\Sigma^{-1}\theta_{\tilde{\ell}} \leq \kappa(\Sigma)\|\theta_{\tilde{\ell}}\|_{E_D}^2.$$

Combining the preceding displays, the proof is now completed as follows:

$$\begin{aligned} \|\theta_{max} - \theta_{\rm init}\|_{E_D}^2 &\lesssim \|\theta_{max}\|_{E_D}^2 + \|\theta_{\rm init}\|_{E_D}^2 \\ &\lesssim \kappa(\Sigma)\|\theta_{\tilde{\ell}}\|_{E_D}^2 + \mathbf{r}^2 + R^2 \lesssim \kappa(\Sigma)\big[1 + \mathbf{r}^2 + R^2\big]. \end{aligned}$$

□

From the previous theorem and concentration inequalities for Markov chains one can obtain the following bound on the computation of posterior statistics by ergodic averages of ϑ_k collected after some burn-in time $J_{in} \in \mathbb{N}$. Specifically, if we define, for any $H : E_D \to \mathbb{R}$ integrable with respect to $\Pi(\cdot|D_N)$, the random variable

$$\hat{\pi}^J_{J_{in}}(H) = \frac{1}{J}\sum_{k=J_{in}+1}^{J_{in}+J} H(\vartheta_k), \tag{5.28}$$

we obtain the following non-asymptotic concentration bound, where $\|H\|_{Lip}$ denotes the Lipschitz constant $\sup_{\theta\neq\theta'\in E_D}[|H(\theta)-H(\theta')|/\|\theta-\theta'\|_{E_D}]$ of H.

Theorem 5.2.2 (Lipschitz functionals)**.** *In the setting of the previous theorem, there exist further constants $c_3, c_4 > 0$ such that for any $\rho > 0$, any burn-in period*

$$J_{in} \geq \frac{c_3}{m\gamma} \times \log\Big(1 + \frac{1}{\rho + b(\gamma)} + \tau(\Sigma, R) + \frac{D}{m}\Big), \tag{5.29}$$

any $J \in \mathbb{N}$, any Lipschitz function $H : \mathbb{R}^D \to \mathbb{R}$, any $t \geq \sqrt{8}\|H\|_{Lip}\sqrt{\rho + b(\gamma)}$ and on the event $\mathcal{E}$ from (5.24), we have

$$\mathbf{P}_{\theta_{\text{init}}}\Big(\big|\hat{\pi}^J_{J_{in}}(H) - E^{\Pi}[H|D_N]\big| \geq t\Big) \leq 2\exp\Big(-c_4\frac{t^2m^2J\gamma}{\|H\|^2_{Lip}(1+1/(mJ\gamma))}\Big). \tag{5.30}$$

Proof. For any $t \geq 0$ and any Lipschitz function $H : E_D \to \mathbb{R}$ we have

$$\begin{aligned} &\mathbf{P}_{\theta_{\text{init}}}\Big(\big|\hat{\pi}^J_{J_{in}}(H) - E^{\Pi}[H|D_N]\big| \geq t\Big) \\ &\quad\leq \mathbf{P}_{\theta_{\text{init}}}\Big(\big|\hat{\pi}^J_{J_{in}}(H) - \mathbf{E}_{\theta_{\text{init}}}[\hat{\pi}^J_{J_{in}}(H)]\big| \geq t - \big|\mathbf{E}_{\theta_{\text{init}}}[\hat{\pi}^J_{J_{in}}(H)] - E^{\Pi}[H|D_N]\big|\Big). \end{aligned} \tag{5.31}$$

To further estimate the right side, note that for c_3 large enough and any $k \geq J_{in}$, by (5.29) and Theorem 5.2.1, we have

$$W_2^2(\mathcal{L}(\vartheta_k), \Pi(\cdot|D_N)) \leq 2(\rho + b(\gamma)).$$

Noting that (6.53) below in fact holds for any probability measure μ and thus in particular for $\mu = \Pi(\cdot|D_N)$, it follows that for any Lipschitz function $H : E_D \to \mathbb{R}$,

$$\big(\mathbf{E}_{\theta_{\text{init}}}[\hat{\pi}^J_{J_{in}}(H)] - E^{\Pi}[H|D_N]\big)^2 \leq 2\|H\|^2_{Lip}(\rho + b(\gamma)).$$

Thus if $t \geq 0$ satisfies our hypothesis, then applying Proposition 6.2.7 to both H and $-H$ yields that the r.h.s. in (5.31) is further bounded by

$$\mathbf{P}_{\theta_{\text{init}}}\Big(\big|\hat{\pi}^J_{J_{in}}(H) - \mathbf{E}_{\theta_{\text{init}}}[\hat{\pi}^J_{J_{in}}(H)]\big| \geq t/2\Big) \leq 2\exp\Big(-c\frac{t^2m^2J\gamma}{\|H\|^2_{Lip}(1+1/(mJ\gamma))}\Big).$$

$\square$

From the last theorem one can obtain as a direct consequence the following guarantee for computation of the posterior mean $E^\Pi[\theta|D_N]$ by the ergodic average accrued along the Markov chain.

Corollary 5.2.3. *In the setting of Theorem 5.2.2, if we define the ergodic average*

$$\bar\theta^J_{J_{in}} = \frac{1}{J}\sum_{k=J_{in}+1}^{J_{in}+J} \vartheta_k$$

then on the event $\mathcal{E}$ from (5.24) and for $t \geq \sqrt{8}\sqrt{\rho + b(\gamma)}$, we have for some constant $c_5 > 0$ that

$$\mathbf{P}_{\theta_{\text{init}}}\Big(\big\|\bar\theta^J_{J_{in}} - E^\Pi[\theta|D_N]\big\|_{E_D} \geq t\Big) \leq 2D\exp\Big(-c_5\frac{t^2m^2J\gamma}{D(1+1/(mJ\gamma))}\Big). \tag{5.32}$$

Proof. We first estimate the probability to be bounded by

$$\mathbf{P}_{\theta_{\text{init}}}\Big(\big\|\bar\theta^J_{J_{in}} - \mathbf{E}_{\theta_{\text{init}}}\big[\bar\theta^J_{J_{in}}\big]\big\|_{E_D} \geq t - \big\|\mathbf{E}_{\theta_{\text{init}}}\big[\bar\theta^J_{J_{in}}\big] - E^\Pi[\theta|D_N]\big\|_{E_D}\Big).$$

Next, for any $k \geq 1$, let ν_k denote an optimal coupling between $\mathcal{L}(\vartheta_k)$ and $\Pi(\cdot|D_N)$ (cf. Theorem 4.1 in [133]). Then by Jensen's inequality and the definition of W_2 from (5.15),

$$\begin{aligned}
\big\|\mathbf{E}_{\theta_{\text{init}}}\big[\bar\theta^J_{J_{in}}\big] - E^\Pi[\theta|D_N]\big\|^2_{E_D} &= \Big\|\frac{1}{J}\sum_{k=J_{in}+1}^{J_{in}+J}\int_{E_D\times E_D}(\theta-\theta')d\nu_k(\theta,\theta')\Big\|^2_{E_D} \\
&= \sum_{j=1}^{D}\Big(\frac{1}{J}\sum_{k=J_{in}+1}^{J_{in}+J}\int_{E_D\times E_D}(\theta_j-\theta'_j)d\nu_k(\theta,\theta')\Big)^2 \\
&\leq \frac{1}{J}\sum_{k=J_{in}+1}^{J_{in}+J}\int_{E_D\times E_D}\sum_{j=1}^{D}(\theta_j-\theta'_j)^2 d\nu_k(\theta,\theta') \\
&= \frac{1}{J}\sum_{k=J_{in}+1}^{J_{in}+J} W_2^2(\mathcal{L}(\vartheta_k), \Pi[\cdot|D_N]).
\end{aligned}$$

Thus we obtain from (5.26), (5.29) (as after (5.31)) that

$$\big\|E_{\theta_{\text{init}}}\big[\bar\theta^J_{J_{in}}\big] - E^\Pi[\theta|D_N]\big\|_{E_D} \leq \sqrt{2}\sqrt{\rho + b(\gamma)}.$$

Now for any $j = 1, ..., d$, let us write $H_j : E_D \to \mathbb{R}$, $\theta \mapsto \theta_j$, for the j-the coordinate projection map, of Lipschitz constant 1. Then in the notation (5.28) we can write

$$[\bar\theta^J_{J_{in}}]_j = \hat\pi^J_{J_{in}}(H_j), \qquad j = 1, ..., D.$$

For $t \geq \sqrt{8(\rho + b(\gamma))}$ and applying Proposition 6.2.7 as in the proof of Theorem 5.2.2 as well as a union bound gives

$$\begin{aligned}
\mathbf{P}_{\theta_{\text{init}}}\Big(\big\|\bar{\theta}^J_{J_{in}} - E^\Pi[\theta|D_N]\big\|_{E_D} \geq t\Big) &\leq \mathbf{P}_{\theta_{\text{init}}}\Big(\big\|\bar{\theta}^J_{J_{in}} - \mathbf{E}_{\theta_{\text{init}}}\big[\bar{\theta}^J_{J_{in}}\big]\big\|_{E_D} \geq t/2\Big) \\
&= \mathbf{P}_{\theta_{\text{init}}}\Big(\sum_{j=1}^{D}\Big[\hat{\pi}^J_{J_{in}}(H_j) - \mathbf{E}_{\theta_{\text{init}}}\big[\hat{\pi}^J_{J_{in}}(H_j)\big]\Big]^2 \geq \frac{t^2}{4}\Big) \\
&\leq \sum_{j=1}^{D}\mathbf{P}_{\theta_{\text{init}}}\Big(\Big[\hat{\pi}^J_{J_{in}}(H_j) - \mathbf{E}_{\theta_{\text{init}}}\big[\hat{\pi}^J_{J_{in}}(H_j)\big]\Big]^2 \geq \frac{t^2}{4D}\Big) \\
&\leq 2D\exp\Big(-c\frac{t^2m^2J\gamma}{D\big[1 + 1/(mJ\gamma)\big]}\Big).
\end{aligned}$$

□

The two previous results imply that one can compute the posterior mean (or posterior functionals $E^\Pi[H|D_N]$ with $\|H\|_{Lip} \leq 1$) within precision $\varepsilon > 0$ as long as $\varepsilon \gtrsim \sqrt{\rho}$: For instance if γ is chosen as

$$\gamma \simeq \min\Big\{\frac{\varepsilon^2 m^2}{D\Lambda^2}, \frac{\varepsilon m^{3/2}}{D^{1/2}\Lambda^2}\Big\}, \tag{5.33}$$

then the overall number of required MCMC iterations $J_{in} + J$ depends polynomially on the quantities $N, D, m^{-1}, \Lambda, \varepsilon^{-1}$. When the latter three constants also exhibit at most polynomial growth in N, D, we can deduce that polynomial-time computation of such posterior characteristics is feasible, on the event $\mathcal{E}$ from (5.24) at computational cost

$$J_{in} + J = O(N^{b_1}D^{b_2}\varepsilon^{-b_3}), \quad b_1, b_2, b_3 > 0,$$

with $\mathbf{P}_{\theta_{\text{init}}}$-probability as close to one as desired.

5.2.2 On failure of 'cold start' MCMC in high dimensions

In Subsection 5.3.2 to follow we will use the preceding theorems to show that properly initialised Langevin-type MCMC methods can compute posterior characteristics in Darcy's inverse problem in a run-time that scales at most *polynomially* in the key parameters D, N and the desired precision level. This is a non-trivial result in view of the widely accepted 'folklore' that sampling from high-dimensional target measures is 'exponentially hard' in absence of strong structural assumptions: for instance, a Markov chain may require a long time to 'escape' from a local optimum of the posterior density, and if such 'energy wells' have depth N, the algorithmic run-time towards the invariant measure scales exponentially in N. Also, the mixing time of Markov chains for convergence to equilibrium may scale exponentially in dimension D without further special assumptions.

In this subsection we give a concrete construction of such a 'computational hardness phenomenon' in a context with Gaussian process priors and average 'population' log-likelihoods that are not globally log-concave. Specifically we consider sampling from a probability measure

$$d\bar{\Pi}(\theta) \propto e^{-Nw(\|\theta\|_{\mathbb{R}^D})} d\Pi(\theta), \quad \theta \in \mathbb{R}^D, \tag{5.34}$$

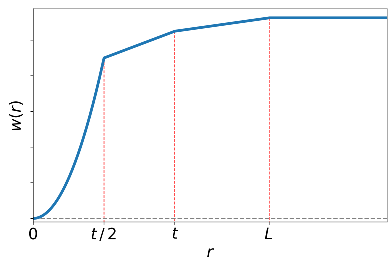
that arises from a 'radial', expected log-likelihood $\ell(\theta) = NE_{\theta_0}\ell_N(\theta) = -Nw(\|\theta\|_{\mathbb{R}^D})$ with scalar function w to be specified below, and a centred Gaussian prior $\Pi = Law(\theta)$ on $\mathbb{R}^D$ given by

$$\theta = (\theta_j)_{j=1}^D, \ \theta_j = j^{-\alpha/d} g_j, \quad g_j \sim^{iid} N(0,1), \quad \alpha > d/2. \tag{5.35}$$

In view of (6.14) this is akin to the prior used in (5.4) but without the N-dependent 'rescaling'. The latter could be considered as well in the results that follow (by just increasing the constant b to be introduced by $(d/2)/(2\alpha + d)$ in the proofs). The results presented further extend without difficulty (see [11]) to the 'empirical' versions ℓ_N replacing $N\ell(\theta)$ but to expedite proofs we consider the 'average' case only, which is sufficient to illustrate the main ideas. We will also assume $\theta_0 = 0$ so that the prior is already centred at the true parameter.

To construct w, let us set $b = (\alpha/d) - (1/2) > 0$ for α the regularity level from (5.35). Then define for any $r \geq 0$,

$$\begin{aligned} w(r) = {} & 4(Tr)^2 1_{[0,t/2)}(r) \\ & + [(Tt)^2 + T(r - t/2)] 1_{[t/2,t)}(r) \\ & + [(Tt)^2 + (Tt/2) + \rho(r - t)] 1_{[t,L)}(r) \\ & + [(Tt)^2 + (Tt/2) + \rho(L - t)] 1_{[L,\infty)}(r), \end{aligned}$$



where, for any choice of $\rho \in (0, 1]$ and $0 < t_b < 1/2$, we set

$$t = t_b N^{-b}, \ T = T_b N^b, \tag{5.36}$$

with $T_b > 0$ a fixed constant to be chosen, and arbitrary fixed $L > t$. Note that such w is 'unimodal' around 0 and the resulting Gibbs measure $\bar{\Pi}$ is locally (but not globally) log-concave near its unique maximum at $\theta = 0$; the density of $\bar{\Pi}$ is strictly decreasing along each half-line departing from the origin.

The theorem that follows will imply that in high-dimensional models with $D \to \infty$, a badly initialised 'local' MCMC algorithm may never (in 'finite time') visit the region where the target measure $\bar{\Pi}$ places the bulk of its mass, i.e., where it is statistically informative in the sense that the 'log-likelihood' term $\ell(\theta)$ dominates the prior. As a consequence the inference obtained is no better than a random number generator that ignores the observed data. The hardness of sampling from $\bar{\Pi}$ arises here not from a 'well' (local optimum) but from a combination of high dimensionality and lack of 'signal in the tails'. We refer to [11] for a more detailed evaluation of these negative results. They showcase the need to exploit specific structure of non-linear models when studying algorithms in high-dimensions (as we will do in Section 5.3).

To formulate the main result of this subsection, let us define balls

$$B_r = \{\theta \in \mathbb{R}^D : \|\theta\|_{\mathbb{R}^D} \le r\}, \; r > 0, \tag{5.37}$$

centred at $\theta_0 = 0$, as well as the annuli

$$\Theta_{r,\varepsilon} = \{\theta \in \mathbb{R}^D : \|\theta\|_{\mathbb{R}^D} \in (r, r+\varepsilon)\}, \; \varepsilon > 0. \tag{5.38}$$

We consider MCMC methods given by Markov chains $(\vartheta_k : k \in \mathbb{N})$ that target the probability measure $\bar{\Pi}$ from (5.34), more precisely, that have $\bar{\Pi}$ as their invariant measure. We also assume that their dynamics are 'local' in the sense that the step-sizes from ϑ_k to ϑ_{k+1} at each iterate k cannot exceed a Euclidean distance η.

Theorem 5.2.4. *Consider the probability measure $\bar{\Pi}$ from (5.34) with prior Π on $\mathbb{R}^D$ from (5.35) for some $\alpha > d/2$ and w as in and before (5.36). Assume $D/N \simeq \kappa$ as $N \to \infty$ for some $\kappa > 0$, and set $b = (\alpha/d) - (1/2) > 0$. Then there exist fixed constants $s_b \in (t_b, 1/2)$ and $T_b > 0$ for which the following statements hold true:*

i) *We have $\bar{\Pi}(B_{s_b N^{-b}}) = 1 - O(e^{-N})$ as $N \to \infty$.*

ii) *There exists $\varepsilon, v > 0$ depending on κ, α, d such that for any Markov chain $(\vartheta_k : k \in \mathbb{N})$ of step-size $\eta \in (0, s_b N^{-b})$ and with invariant measure $\bar{\Pi}$, we can find an initialisation point $\vartheta_0 \in \Theta_{N^{-b},\varepsilon N^{-b}}$ for which the hitting time τ_{B_s} for ϑ_k to reach $B_{s_b N^{-b}}$ is lower bounded as*

$$\tau_{B_s} = \inf\{k : \vartheta_k \in B_{s_b N^{-b}}\} \ge \exp\{vD\}$$

with probability approaching one under the law of the Markov chain.

The proof will show that the result holds for ϑ_0 drawn at random from an absolutely continuous distribution on $\Theta_{N^{-b},\varepsilon N^{-b}}$, and then employs the probabilistic method to deduce existence of a fixed 'worst case' initialiser.

The theorem is stated for Markov chains with 'deterministic' step-sizes, whereas many concrete MCMC methods such as ULA have *random* step-sizes, that however do not exceed η with high probability under the randomisation scheme of the algorithm. The previous theorem *does* extend to such settings, and in particular to methods such as pCN, MALA – as is shown in [11] – but we will only prove the above version here for simplicity. Note that while the step sizes have to be sufficiently small, they are of polynomial order in $1/N$, wheras the hitting time lower bound scales exponentially in $N \simeq D$.

Theorem 5.2.4 does require non-degenerate asymptotics $D/N \simeq \kappa > 0$, to permit 'volumetric' arguments and 'unimodal' target measures $\bar{\Pi}$. If in contrast $\bar{\Pi}$ has a 'well' (i.e., w has a local optimum away from 0), then the ideas underlying the proof that follows can be adapted to give an exponential lower bound $\tau_{B_s} \geq e^{vN}$ with high probability also when $D/N \to 0$.

Proof of Theorem 5.2.4

We first show **Part ii)**, which relies on the following 'bottleneck' argument for Markov chains, essentially due to [71]. The result is general but we present it in our setting for (ϑ_k) moving between annuli from (5.38) and targeting $\bar{\Pi}$ from (5.34). Notice that we always have $\bar{\Pi}(\Theta_{\sigma,\epsilon}) > 0$ in this case.

Proposition 5.2.5. *Consider a Markov chain $(\vartheta_k : k \in \mathbb{N})$ on $\mathbb{R}^D$ with invariant measure $\bar{\Pi}$ and such that*

$$\frac{\bar{\Pi}(\Theta_{s,\eta})}{\bar{\Pi}(\Theta_{\sigma,\epsilon})} \leq e^{-\nu N} \tag{5.39}$$

for some $\nu, \epsilon > 0$ and $0 < s < \sigma, \eta < \sigma - s$. Suppose ϑ_0 is started in $\Theta_{\sigma,\epsilon}$, drawn from the conditional distribution $\bar{\Pi}(\cdot|\Theta_{\sigma,\epsilon})$, and denote by τ_s the hitting time of the Markov chain onto $\Theta_{s,\eta}$, that is, the number τ_s of iterates required until ϑ_k visits the set $\Theta_{s,\eta}$. Then

$$\Pr(\tau_s \leq K) \leq K e^{-\nu N}, \quad K > 0.$$

Proof. The probability we wish to bound equals

$$\begin{aligned}
&\Pr(\vartheta_k \in \Theta_{s,\eta} \text{ for some } 1 \leq k \leq K | \vartheta_0 \in \Theta_{\sigma,\epsilon}) \\
&= \frac{\Pr(\vartheta_0 \in \Theta_{\sigma,\epsilon}, \vartheta_k \in \Theta_{s,\eta} \text{ for some } 1 \leq k \leq K)}{\bar{\Pi}(\Theta_{\sigma,\epsilon})} \leq \frac{\sum_{k \leq K} \Pr(\vartheta_k \in \Theta_{s,\eta})}{\bar{\Pi}(\Theta_{\sigma,\epsilon})} \\
&\leq K \frac{\bar{\Pi}(\Theta_{s,\eta})}{\bar{\Pi}(\Theta_{\sigma,\epsilon})} \leq K e^{-\nu N}
\end{aligned}$$

since the marginal distributions of ϑ_k are all $\bar{\Pi}$ when $\vartheta_0 \sim \bar{\Pi}$. $\square$

The last proposition holds 'on average' for initialisers $\vartheta_0 \sim \bar{\Pi}(\cdot|\Theta_{\sigma,\epsilon})$. But since $\Pr = E_{\bar{\Pi}(\cdot|\Theta_{\sigma,\epsilon})} \Pr_{\vartheta_0}$ where $\Pr_{\vartheta_0}$ is the law of the Markov chain started at ϑ_0, the hitting time inequality holds at least for one point in $\Theta_{\sigma,\epsilon}$ since $\inf_{\vartheta_0} \Pr_{\vartheta_0} \leq E_{\bar{\Pi}(\cdot|\Theta_{\sigma,\epsilon})} \Pr_{\vartheta_0}$ ('the probabilistic method').

In our setting $\eta < s_b N^{-b} = s$ with $t_b < s_b < 1/2$ and we will apply Proposition 5.2.5 with $\sigma = N^{-b}, \epsilon = \varepsilon N^{-b}$. If we initialise at $\vartheta_0 \in \Theta_{N^{-b},\varepsilon N^{-b}}$, then under our step-size condition the Markov chain needs to visit $\Theta_{s_b N^{-b},\eta}$ in order to reach $B_{s_b N^{-b}}$. Therefore the hitting time for the ball $B_{s_b N^{-b}}$ is lower bounded by the hitting time for the 'intermediate' set $\Theta_{s_b N^{-b},\eta}$, and Theorem 5.2.4ii) with $v = \nu/2$ will follow if we can verify (5.39) for these choices of s, η, σ and some ε, ν. Taking logarithms, the required inequality (5.39) becomes

$$\frac{1}{N} \log \bar{\Pi}(\Theta_{s_b N^{-b},\eta}) - \frac{1}{N} \log \bar{\Pi}(\Theta_{N^{-b},\varepsilon N^{-b}}) \leq -\nu$$

which in view of (5.34) and the choice of w from before (5.36) is implied by

$$\frac{1}{N} \log \frac{\Pi(\Theta_{s_b N^{-b},\eta})}{\Pi(\Theta_{N^{-b},\varepsilon N^{-b}})} \leq -\nu - \rho N^{-b}(1 + \varepsilon - s_b). \tag{5.40}$$

Annuli farther away from the origin will have larger prior volume and in high dimensions $D/N \to \kappa$ this induces a 'free energy barrier', permitting one to verify (5.40), as we now show.

Lemma 5.2.6. *For every $s_b > t_b$ we can choose $\nu, \varepsilon > 0$ depending only on s_b, κ, α, d such that (5.40) holds.*

Proof. Lemma 5.2.7 and the hypotheses on η imply

$$\begin{aligned}\Pi(\Theta_{s_b N^{-b},\eta}) &= \Pi\big(\|\theta\|_{\mathbb{R}^D} \in (s_b N^{-b}, s_b N^{-b} + \eta)\big) \leq \Pi\big(\|\theta\|_{\mathbb{R}^D} \leq 2 s_b N^{-b}\big) \\ &\leq e^{-c_0 N (2s_b + \kappa^{-\alpha/d}(2s_b)^{-\tau/2})^{-\tau}}.\end{aligned}$$

To lower bound $\Pi(\Theta_{N^{-b},(1+\varepsilon)N^{-b}})$, we choose ε large enough such that

$$\bar{c}_0(1+\varepsilon)^{-\tau} < c_0(1 + \kappa^{-\alpha/d})^{-\tau},$$

so that for all N large enough

$$\begin{aligned}\Pi\big(\|\theta\|_{\mathbb{R}^D} \in (N^{-b}, (1+\varepsilon)N^{-b})\big) &= \Pi(\|\theta\|_{\mathbb{R}^D} \leq (1+\varepsilon)N^{-b}) - \Pi(\|\theta\|_{\mathbb{R}^D} \leq N^{-b})) \\ &\geq e^{-\bar{c}_0(1+\varepsilon)^{-\tau} N} - e^{-c_0(1+\kappa^{-\alpha/d})^{-\tau} N} \\ &\geq e^{-2\bar{c}_0(1+\varepsilon)^{-\tau} N}.\end{aligned} \tag{5.41}$$

The required bound (5.40) will thus follow if

$$c_0(2s_b + \kappa^{-\alpha/d}(2s_b)^{-\tau/2})^{-\tau} \geq 2\bar{c}_0(1+\varepsilon)^{-\tau} + \nu + \rho N^{-b}(1 + \varepsilon - s_b). \tag{5.42}$$

If we define ν to equal to 1/2 of the l.h.s. of the last display then (5.42) will follow for any $s_b \leq 1/2, \rho \leq 1$ and the given κ, α, d by choosing ε large enough and whenever N is large enough. $\square$

To complete the proof of the theorem we now show **Part i)** for some fixed choice of $t_b < s_b < 1/2$ and T_b: Let us write

$$\mathbb{G}(A) := \int_A e^{-Nw(\|\theta\|_{\mathbb{R}^D})} d\Pi(\theta)$$

for any measurable subset A of $\mathbb{R}^D$. Using Lemma 5.2.7, we have

$$\frac{1}{N} \log \mathbb{G}(B_{t_b N^{-b}/2}) \geq -w(t_b N^{-b}/2) - \bar{c}_0 \big(\frac{t_b}{2}\big)^{-\tau}$$

as well as

$$\frac{1}{N} \log \mathbb{G}(B^c_{s_b N^{-b}}) \leq -w(s_b N^{-b}) + \frac{1}{N} \log \Pi(B^c_{s_b N^{-b}}).$$

Now $\Pi(B^c_{s_b N^{-b}}) \to 1$ by Lemma 5.2.7 and hence by choice of w from before (5.36)

$$\begin{aligned}
&\frac{1}{N} \log \frac{\mathbb{G}(B_{t_b N^{-b}/2})}{\mathbb{G}(B^c_{s_b N^{-b}})} \\
&\geq -(T_b t_b)^2 - c_0 \big(\frac{t_b}{2}\big)^{-\tau} + [(T_b t_b)^2 + (T_b t_b/2) + \rho N^{-b}(s_b - t_b)] + \frac{1}{N} \log \Pi(B^c_s) \\
&\geq \frac{T_b t_b}{2} - c_0 \big(\frac{t_b}{2}\big)^{-\tau} + o(1),
\end{aligned} \tag{5.43}$$

as $N \to \infty$. Now, for $t_b < s_b$ fixed we can choose T_b large enough such that the last quantity exceeds 1. Therefore

$$\frac{\mathbb{G}(B_{t_b N^{-b}/2})}{\mathbb{G}(B^c_{s_b N^{-b}})} \geq e^N \times (1 + o(1)). \tag{5.44}$$

For $M_N = \{\theta : t_b N^{-b}/2 < \|\theta\|_{\mathbb{R}^D} \leq s_b N^{-b}\}$ this further implies

$$Z \equiv \frac{\mathbb{G}(B_{t_b N^{-b}/2}) + \mathbb{G}(M_N)}{\mathbb{G}(B^c_{s_b N^{-b}})} \geq e^N \times (1 + o(1)),$$

and **Part i)** then follows from this bound and

$$\begin{aligned}
\bar{\Pi}(B_{s_b N^{-b}}) &= \frac{\mathbb{G}(B_{t_b N^{-b}/2}) + \mathbb{G}(M_N)}{\mathbb{G}(B_{t_b N^{-b}/2}) + \mathbb{G}(M_N) + \mathbb{G}(B^c_{s_b N^{-b}})} \\
&= \frac{\mathbb{G}(B_{t_b N^{-b}/2}) + \mathbb{G}(M_N)}{(\mathbb{G}(B_{t_b N^{-b}/2}) + \mathbb{G}(M_N))\big(1 + Z^{-1}\big)} = \frac{1}{1 + Z^{-1}}.
\end{aligned}$$

We conclude this section with the proof of the following key volumetric lemma on the small-deviation landscape of ellipsoidally supported high-dimensional Gaussian measures.

Lemma 5.2.7. *Let* Π *be the Gaussian prior on* $\mathbb{R}^D$ *arising from (5.35). Fix* $z > 0$ *and* $\kappa > 0$*, and set*

$$b = \frac{\alpha}{d} - \frac{1}{2}, \ \tau = \frac{1}{b} = \frac{2d}{2\alpha - d}.$$

Then if $D/N \simeq \kappa > 0$*, there exist constants* $\bar{c}_0 > c_0$ *(depending on* b, κ*) such that for all* N $(\geq N_0(z, b))$ *large enough:*

$$c_0(z + \kappa^{-\alpha/d} z^{-\tau/2})^{-\tau} \leq -\frac{1}{N} \log \Pi(\|\theta\|_{\mathbb{R}^D} \leq zN^{-b}) \leq \bar{c}_0 z^{-\tau}. \tag{5.45}$$

Proof. Consider first the 'untruncated prior' (5.35) with series extending to infinity, convergent almost surely in the space ℓ_2 with norm $\|\theta\|_{\ell_2}^2 = \sum_{j=1}^\infty \theta_j^2$, since $\alpha > d/2$. Using Proposition 6.1.1, Remark 6.1.2 and Theorem 6.2.1 with $a = \frac{2d}{2\alpha - d}$, Banach space $B = \ell_2$ and RKHS

$$\mathcal{H} = h^\alpha := \{\theta : \|\theta\|_{h^\alpha}^2 = \sum_j j^{2(\alpha/d)} \theta_j^2 < \infty\},$$

we deduce the two-sided estimate

$$-\log \Pi(\|\theta\|_{\ell_2} \leq \gamma) \simeq \gamma^{-\frac{2d}{2\alpha - d}} = \gamma^{-\tau}, \quad \gamma \to 0. \tag{5.46}$$

Here, restricting to $\gamma \in (0, 1)$, the two-sided equivalence constants depend only on α, d. Setting

$$\gamma = zN^{-b}, \ z > 0, \tag{5.47}$$

and noting that $b\tau = 1$, we hence obtain that for some constants $c_l, c_u > 0$,

$$e^{-c_l z^{-\tau} N} \leq \Pi(\|\theta\|_{\ell_2} \leq zN^{-b}) \leq e^{-c_u z^{-\tau} N}, \quad \text{any } z > 0. \tag{5.48}$$

For the projected prior on $\mathbb{R}^D$, we have by set inclusion and projection,

$$\Pi(\|\theta\|_{\mathbb{R}^D} \leq zN^{-b}) \geq \Pi(\|\theta\|_{L^2} \leq zN^{-b})$$

and hence it only remains to show the first inequality in eq. (5.45). The Gaussian isoperimetric theorem (Theorem 2.6.12 in [61]) and (5.48) imply that for $m \geq 4\sqrt{c_l}$ and some $c > 0$, we have (with Φ denoting the c.d.f. for $N(0, 1)$)

$$\begin{aligned}
\Pi\big(\theta = \theta_1 + \theta_2, \|\theta_1\|_{\ell_2} \leq zN^{-b}, \|\theta_2\|_{h^\alpha} \leq mz^{-\tau/2}\sqrt{N}\big) \\
\geq \Phi\big(\Phi^{-1}\big(\Pi(\{\theta : \|\theta\|_{\ell_2} \leq zN^{-b}\})\big) + mz^{-\tau/2}\sqrt{N}\big) \\
\geq \Phi\big(-\sqrt{2c_l} z^{-\tau/2}\sqrt{N} + mz^{-\tau/2}\sqrt{N}\big) \geq 1 - e^{-cz^{-\tau}N}.
\end{aligned}$$

Then if the event in the last probability is denoted by I we have

$$\Pi(\|\theta_D\|_{\mathbb{R}^D} \leq zN^{-b}) \leq \Pi(\|\theta_D\|_{\mathbb{R}^D} \leq zN^{-b}, I) + e^{-cz^{-\tau}N}.$$

On I, if $D/N \to \kappa > 0$ and by the usual tail estimate for vectors $\theta_2 \in h^\alpha$, we have for some $c' > 0$ the bound

$$\|\theta - \theta_D\|_{\ell_2} \leq \|\theta_1\|_{\ell_2} + c' D^{-\alpha/d} z^{-\tau/2}\sqrt{N} \leq zN^{-b} + c'\kappa^{-\alpha/d} z^{-\tau/2} N^{-b}$$

so that for any $z > 0$,

$$\begin{aligned}\Pi(\|\theta_D\|_{\mathbb{R}^D} \leq zN^{-b}) &\leq \Pi(\|\theta\|_{\ell_2} \leq zN^{-b} + \|\theta - \theta_D\|_{\ell_2}, I) + e^{-cz^{-\tau}N} \\ &\leq \Pi(\|\theta\|_{\ell_2} \leq (2z + c'\kappa^{-\alpha/d} z^{-\tau/2})N^{-b}) + e^{-cz^{-\tau}N} \\ &\leq e^{-c_u(2z + c'\kappa^{-\alpha/d} z^{-\tau/2})^{-\tau}N} + e^{-cz^{-\tau}N},\end{aligned}$$

and hence the lemma follows by appropriate choice of $c_0 > 0$. □

5.3 Application to PDE models

We now explain how the ideas from this chapter apply to Darcy's problem, that is, the PDE (1.2) with diffusion operator $\mathscr{L}_{f_\theta}$ from (1.3) resulting in the forward map $\mathscr{G}$ from (2.8). The theory for the Schrödinger equation is similar and discussed in the notes (see also Ex. 5.4.5) along with further relevant examples.

We first need to derive some analytical properties of the forward map $\mathscr{G} : \Theta \to L^2_\lambda(\mathcal{X})$ from (2.8) restricted to the finite-dimensional approximation space E_D from (5.3) arising as the eigen-spaces E_D of the Dirichlet-Laplacian on $\mathcal{X}$. As ambient parameter space Θ we take $\tilde{H}^\alpha(\mathcal{X}) \subset H^\alpha_0(\mathcal{X})$ from (6.15) for some $\alpha > 0$ to be specified – these spaces contain E_D for all D.

The key result will be Theorem 5.3.2 which verifies the average local curvature condition (3.34) for this forward map on E_D. It is based on the ideas from Section 3.2 and the observation that the stability condition (3.30) for $D\mathscr{G}_\theta$ indeed holds for $\theta \in \Theta$ by virtue of a) the isomorphism properties of general elliptic operators between Sobolev spaces and their characterisation in the basis spanning E_D and b) the stability Lemma 2.1.6. Note that all $\theta \in E_D \subset \tilde{H}^\alpha(\mathcal{X})$ vanish at the boundary $\partial\mathcal{X}$, as is required to apply Lemma 2.1.6. In the present context this is essential as without this restriction $D\mathscr{G}_\theta$ may fail to be injective (recall Ex. 3.4.5). Our construction avoids the counter-examples from Section 4.2.2 – for the 'non-asymptotic' theorems here we never need to consider the L^2-*completion* of our tangent (or parameter) space Θ, nor does the range of the adjoint of $D\mathscr{G}_\theta$ play a role, as was the case in the 'asymptotic' Theorems 3.1.5, 4.2.2 and 4.2.3.

5.3.1 Darcy's problem on the eigen-spaces of the Laplacian

To check Conditions 3.2.1 and 3.2.2 we require bounds on the gradient and Hessian of the real-valued map $\mathscr{G}^x(\theta) \equiv \mathscr{G}(\theta)(x)$ with respect to the variable $\theta \in E_D \simeq \mathbb{R}^D$ (for fixed $x \in \mathcal{X}$).

We have already computed the L^2_λ-linearisation $D\mathscr{G}_{\theta_0}$ of $\mathscr{G}$ at θ_0 in Theorem 3.3.2, and the proof there can be upgraded to $\|\cdot\|_\infty$-differentiability if the perturbations h are bounded in sufficiently strong Sobolev norms. We restrict to $d \le 3$ for simplicity – this permits a straightforward application of the Sobolev-imbedding theorem $H^2 \subset L^\infty$ from (6.4), removing some technicalities in the proof of the general case.

Proposition 5.3.1. *Let $d \le 3$. For $\mathscr{G}$ from (2.8) and $D\mathscr{G}_\theta$ from (3.57) let θ, h belong to $\Theta = \tilde{H}^\alpha, \alpha > 1 + d/2$. Then we have as $\|h\|_{H^\alpha} \to 0$ that*

$$\|\mathscr{G}(\theta + h) - \mathscr{G}(\theta) - D\mathscr{G}_\theta[h]\|_\infty = O(\|h\|^2_{H^\alpha}).$$

In particular, the map $\mathscr{G}^x = \mathscr{G}(\cdot)(x) : E_D \to \mathbb{R}$ has gradient

$$v^T \nabla \mathscr{G}^x(\theta) = -\mathscr{L}^{-1}_{f_\theta}[\nabla \cdot (e^\theta v \nabla u_{f_\theta})](x), \quad x \in \mathcal{X}, \ v \in E_D,$$

where $\mathscr{L}^{-1}_{f_\theta}$ is the inverse of the operator $\mathscr{L}_{f_\theta}$ from (1.3) for Dirichlet boundary conditions (i.e., equal to $L^{-1}_{\gamma,V}$ given in (6.10) with $\gamma = f_\theta, V = 0$).

Proof. Arguing just as in the proof of Theorem 3.3.2 and using the Sobolev imbedding $H^2 \subset L^\infty$ with $d \le 3$ (see (6.4)) we obtain

$$\begin{aligned}\|\mathscr{G}(\theta + h) - \mathscr{G}(\theta) - D\mathscr{G}_\theta[h]\|_\infty &\lesssim \|\mathscr{L}^{-1}_{f_{\theta+h}}[R_1(h)]\|_{H^2} + \|\mathscr{L}^{-1}_{f_{\theta+h}}[R_2(h)]\|_{H^2} \\ &\lesssim \|R_1(h)\|_{L^2} + \|R_2(h)\|_{L^2}\end{aligned}$$

where we have used (6.35) and that $f_\theta, f_{\theta+h}$ are bounded in C^1 again by the Sobolev imbedding. The r.h.s. is shown to be $O(\|h\|^2_{C^1}) = O(\|h\|^2_{H^\alpha})$ by appropriate modifications of the arguments from the proof of Theorem 3.3.2, see Ex. 5.4.3 for details. The formula for the gradient then follows directly since $v \in E_D \subset H^\alpha$. □

The previous proof can be iterated (using also the chain and product rule) to show that $\mathscr{G}^x$ is in fact smooth on E_D. One obtains the formula for 'second derivatives' $D^2\mathscr{G}_\theta$ acting as bilinear forms

$$\begin{aligned}D^2\mathscr{G}_\theta[v_1, v_2] = &-\mathscr{L}^{-1}_{f_\theta}[\nabla \cdot (e^\theta v_1 v_2 \nabla u_{f_\theta})] \\ &+ \mathscr{L}^{-1}_{f_\theta}\Big[\nabla \cdot \big(e^\theta v_1 \nabla \mathscr{L}^{-1}_{f_\theta}[\nabla \cdot (e^\theta v_2 \nabla u_{f_\theta})]\big)\Big] \\ &+ \mathscr{L}^{-1}_{f_\theta}\Big[\nabla \cdot \big(e^\theta v_2 \nabla \mathscr{L}^{-1}_{f_\theta}[\nabla \cdot (e^\theta v_1 \nabla u_{f_\theta})]\big)\Big]\end{aligned}$$

for $v_1, v_2 \in E_D$. In particular for $v \in E_D$ and $x \in \mathcal{X}$ we have

$$\begin{aligned}v^T \nabla^2 \mathscr{G}^x(\theta) v = &-\mathscr{L}^{-1}_{f_\theta}[\nabla \cdot (e^\theta v^2 \nabla u_{f_\theta})](x) \\ &+ 2\mathscr{L}^{-1}_{f_\theta}\Big[\nabla \cdot \big(e^\theta v \nabla \mathscr{L}^{-1}_{f_\theta}[\nabla \cdot (e^\theta v \nabla u_{f_\theta})]\big)\Big](x) \end{aligned} \tag{5.49}$$

The details are only of a technical nature and left to Ex. 5.4.3.

Average curvature for Darcy's problem on E_D

The following lemma checks the local curvature hypothesis (3.34) for the forward map from (2.8) if the natural 'identifiability' condition (2.16) is satisfied – cf. also (2.20) for sufficient conditions on g, h ensuring (2.16). A result comparable to the next theorem holds for the Schrödinger model (2.5) as well – see Ex. 5.4.5 – and also for other PDE models discussed in the notes to this section.

To concentrate on the main ideas we assume that θ_0 is sufficiently regular in a Sobolev sense and do not attempt to optimise the powers of D in the choice of radius $\mathbf{r}$ of $\mathcal{B}_\mathbf{r}$ or in (5.50). Neither do we make any claims about the optimality of the hypothesis $\alpha \geq 5$.

Theorem 5.3.2. *Let $\ell(\theta) = \ell(\theta, (Y, X))$ be as in (3.33) with $\mathscr{G} : E_D \to \mathbb{R}$ from (2.8), $d \leq 3$, with $E_D \subset \Theta$ as in (5.3) for $\alpha \geq 5$, and let $\mathcal{B}_\mathbf{r}$ be as in (5.2) with $\mathbf{r} = rD^{-w}$ where $w = 8/d$ and $r > 0$. Suppose (2.16) holds for all $\theta \in \mathcal{B}_\mathbf{r}$. Let $\theta_0 \in \tilde{H}^5(\mathcal{X})$ satisfy $\|\theta_0\|_{\tilde{H}^5} \leq S$ for some $S > 0$. Then there exist constants $0 < r_S \leq 1, c_1, c_2 > 0$ such that if also $\|\mathscr{G}(\theta_0) - \mathscr{G}(\theta_{0,D})\|_{L^2_\lambda} \leq c_1 D^{-w}$, then for all $D \in \mathbb{N}$ and all $r \leq r_S$,*

$$\inf_{\theta \in \mathcal{B}_\mathbf{r}} \lambda_{min}\big(E_{\theta_0}\big[-\nabla^2 \ell(\theta)\big]\big) \geq c_2 D^{-6/d}. \tag{5.50}$$

Proof. As a preliminary remark we record here that for $\theta \in \mathcal{B}_\mathbf{r}$ we have

$$\|\theta\|_{H^5} \lesssim \|\theta\|_{\tilde{H}^5} \leq \|\theta_{0,D}\|_{\tilde{H}^5} + \|\theta - \theta_{0,D}\|_{\tilde{H}^5} \leq S + D^{5/d} r D^{-w} \lesssim 1 \tag{5.51}$$

where we have used the estimate $\|v\|_{\tilde{H}^5} \lesssim D^{5/d}\|v\|_{E_D}$ for $v \in E_D$, valid by (6.14), (6.15) below. By (6.4) this implies $\|\theta\|_\infty$ is uniformly bounded. Hence by the chain rule we can further restrict to $f_\theta = e^\theta$ bounded in H^5 norm and by Proposition 6.1.5 then also to u_{f_θ} bounded in H^6-norm. By the Sobolev imbedding (6.4) (with $d \leq 3$) this implies

$$\|u_{f_\theta}\|_{C^4} + \|f_\theta\|_{C^3} \leq \|u_{f_\theta}\|_{H^6} + \|f_\theta\|_{H^5} \leq B \tag{5.52}$$

for a constant B that is uniform in $\mathcal{B}_\mathbf{r}$. The preceding estimates will be used tacitly in the proof.

We have from (3.39)

$$-\nabla^2 \ell(\theta) = \nabla \mathscr{G}^X(\theta) \nabla \mathscr{G}^X(\theta)^T - (Y - \mathscr{G}^X(\theta)) \nabla^2 \mathscr{G}^X(\theta).$$

Using this, Proposition 5.3.1 and (5.49), we obtain that for any $v \in E_D$,

$$\begin{aligned} v^T E_{\theta_0}[-\nabla^2 \ell(\theta)] v = {} & \|\mathscr{L}_{f_\theta}^{-1}[\nabla \cdot (e^\theta v \nabla u_{f_\theta})]\|_{L^2_\lambda}^2 \\ & - \big\langle \mathscr{G}(\theta_0) - \mathscr{G}(\theta), 2\mathscr{L}_{f_\theta}^{-1}\big[\nabla \cdot \big(e^\theta v \nabla \mathscr{L}_{f_\theta}^{-1}[\nabla \cdot (e^\theta v \nabla u_{f_\theta})]\big)\big]\big\rangle_{L^2_\lambda} \\ & + \big\langle \mathscr{G}(\theta_0) - \mathscr{G}(\theta), \mathscr{L}_{f_\theta}^{-1}[\nabla \cdot (e^\theta v^2 \nabla u_{f_\theta})]\big\rangle_{L^2_\lambda} \\ =: {} & I - II + III. \end{aligned} \tag{5.53}$$

We lower bound $\lambda_{min}\big(E_{\theta_0}\big[-\nabla^2\ell(\theta)\big]\big)$ by lower bounding the last quantity by a constant multiple times $\|v\|_{E_D}^2$ for all $v \in E_D$ such that $\|v\|_{E_D} = 1$.

a) Lower bound for I: This is the key step. By the interpolation inequality (6.5) for Sobolev norms we have for any $0 \neq w \in H^4(\mathcal{X})$ that

$$\|w\|_{H^2} \lesssim \|w\|_{L^2}^{1/2}\|w\|_{H^4}^{1/2} \Rightarrow \|w\|_{L^2} \gtrsim \frac{\|w\|_{H^2}^2}{\|w\|_{H^4}}. \tag{5.54}$$

We also note that by (6.15), (1.6), (6.21) and the remarks after (6.16) we have

$$\|w\|_{H^2} \simeq \|w\|_{\tilde{H}^2} \simeq \|\mathscr{L}_{f_\theta} w\|_{L^2}, \text{ any } w \in H_0^2(\mathcal{X}).$$

Applying what precedes to $w = \mathscr{L}_{f_\theta}^{-1}[\nabla\cdot(e^\theta v \nabla u_{f_\theta})] \in H_0^2(\mathcal{X})$ with $v \neq 0$ gives, using also Lemma 2.1.6 and Proposition 6.1.6 with $\beta = 3$,

$$\begin{aligned}
\|\mathscr{L}_{f_\theta}^{-1}[\nabla\cdot(e^\theta v\nabla u_{f_\theta})]\|_{L^2} &\gtrsim \frac{\|\nabla\cdot(e^\theta v \nabla u_{f_\theta})\|_{L^2}^2}{\|\nabla\cdot(e^\theta v\nabla u_{f_\theta})\|_{H^2}} \\
&\gtrsim \frac{e^{-2\|\theta\|_\infty}\|v\|_{L^2}^2}{\|e^\theta\|_{C^3}\|u_{f_\theta}\|_{C^4}\|v\|_{H^3}} \\
&\gtrsim D^{-3/d}\|v\|_{L^2}.
\end{aligned}$$

in view of the estimate $\|v\|_{H^3} \lesssim \|v\|_{\tilde{H}^3} \lesssim D^{3/d}\|v\|_{L^2}$ for $v \in E_D \subset \tilde{H}^3 \subset H^3$, valid by (6.15) and (6.14). In conclusion we have proved that for all $0 \neq v \in E_D$,

$$\|\mathscr{L}_{f_\theta}^{-1}[\nabla\cdot(e^\theta v\nabla u_{f_\theta})]\|_{L^2_\lambda}^2 \gtrsim D^{-6/d}\|v\|_{E_D}^2.$$

b) Upper bounds for II, III: For $\theta \in \mathcal{B}_{\mathbf{r}}$ the Cauchy-Schwarz inequality and the local $E_D \to L^2_\lambda$-Lipschitz property of $\mathscr{G}$ (Proposition 2.1.3 with $\kappa = 0$ and (5.51)) imply that the terms II, III are both bounded by $(Lr + c_1)D^{-w}(i + ii)$, and where i is bounded by

$$\begin{aligned}
&\big\|\mathscr{L}_{f_\theta}^{-1}\big[\nabla\cdot\big(e^\theta v\nabla\mathscr{L}_{f_\theta}^{-1}[\nabla\cdot(e^\theta v\nabla u_{f_\theta})]\big)\big]\big\|_{L^2_\lambda} \\
&\lesssim \|e^\theta v\nabla\mathscr{L}_{f_\theta}^{-1}[\nabla\cdot(e^\theta v\nabla u_{f_\theta})]\|_{L^2_\lambda} \\
&\lesssim \|v\|_\infty\|\mathscr{L}_{f_\theta}^{-1}[\nabla\cdot(e^\theta v\nabla u_{f_\theta})]\|_{H^1} \\
&\lesssim \|v\|_\infty\|e^\theta v\nabla u_{f_\theta}\|_{L^2} \leq \|v\|_\infty\|v\|_{L^2}\|f_\theta\|_\infty\|u_{f_\theta}\|_{C^1} \\
&\lesssim D^{(1/2)+\epsilon}\|v\|_{L^2}^2
\end{aligned}$$

using (6.36), (6.8), the divergence theorem (6.7) as well as

$$\|v\|_\infty \lesssim \|v\|_{H^{d(1/2+\epsilon)}} \lesssim D^{(1/2)+\epsilon}\|v\|_{L^2}, v \in E_D,$$

again by (6.15) and (6.14); while ii is bounded by

$$\begin{aligned}\|\mathscr{L}_{f_\theta}^{-1}[\nabla\cdot(e^\theta v^2\nabla u_{f_\theta})]\|_{L^2_\lambda} &\lesssim \|e^\theta v^2 \nabla u_{f_\theta}\|_{L^2}\\ &\leq \|f_\theta\|_\infty \|u_{f_\theta}\|_{C^1}\|v\|_\infty\|v\|_{L^2_\lambda}\\ &\lesssim D^{(1/2)+\epsilon}\|v\|_{L^2}^2\end{aligned}$$

using similar arguments.

To conclude, since $w = 8/d$ and $d < 4$ we obtain the overall bound

$$I - II + III \geq A_1 D^{-6/d}\|v\|_{E_D}^2 - A_2(r+c_1)D^{-w}D^{(1/2)+\epsilon}\|v\|_{E_D}^2 \geq \frac{A_1}{2}D^{-6/d}\|v\|_{E_D}^2$$

for some constants $A_1, A_2 > 0$, all r, c_1 small enough, and all $0 \neq v \in E_D$, proving the theorem. □

Using minor variations of the arguments from the preceding proof one can easily verify the second part of Condition 3.2.2.

Proposition 5.3.3. *Under the hypotheses of Theorem 5.3.2, the bound (3.35) in Condition 3.2.2 holds for $\kappa_1 = 0$.*

Proof. We only consider the operator norm of the expected Hessian, the other terms are of smaller order and bounded in a similar way. We can again use the decomposition (5.53) and the previous proof already showed that $II + III \lesssim D^{-w}D^{1/2+\epsilon} \lesssim 1$. To upper bound I we use (6.36) and (5.52) and the divergence theorem to obtain

$$\sqrt{I} \leq \bar{c}\|\nabla\cdot(e^\theta v \nabla u_{f_\theta})\|_{(H_0^1)^*} \leq \|f_\theta\|_\infty \|u_{f_\theta}\|_{C^1}\|v\|_{L^2} \leq c\|v\|_{E_D}$$

which after collecting terms implies the result. □

Verifying Condition 3.2.1

There are several terms to check when verifying Condition 3.2.1, all following the same standard proof template. We encounter similar terms as in the proof of Theorem 5.3.2, but now need to bound supremum norms instead of L^2_λ-norms, or, by the Sobolev imbedding $H^2 \subset L^\infty$ ($d \leq 3$), corresponding H^2-norms.

For instance for the gradient $\nabla\mathscr{G}^x(\theta)$ from Proposition 5.3.1, we can use (6.35), (5.51), (5.52) for $\theta \in \mathcal{B}_{\mathbf{r}}$ as in Theorem 5.3.2, and (6.15), (6.14) to obtain

$$\begin{aligned}\sup_{x\in\mathcal{X}}|v^T\nabla\mathscr{G}^x(\theta)| &\lesssim \|\mathscr{L}_{f_\theta}^{-1}[\nabla\cdot(e^\theta v\nabla u_{f_\theta})]\|_{H^2} \lesssim \|e^\theta v \nabla u_{f_\theta}\|_{H^1}\\ &\lesssim \|f_\theta\|_{C^1}\|u_{f_\theta}\|_{C^2}\|v\|_{H^1}\\ &\lesssim D^{1/d}\|v\|_{E_D},\end{aligned}$$

while the second term featuring in the Hessian from (5.49) is of the order

$$\begin{aligned}
&\big\|\mathscr{L}_{f_\theta}^{-1}\big[\nabla\cdot\big(e^\theta v\nabla\mathscr{L}_{f_\theta}^{-1}[\nabla\cdot(e^\theta v\nabla u_{f_\theta})]\big)\big]\big\|_{H^2}\\
&\lesssim \|e^\theta v\nabla\mathscr{L}_{f_\theta}^{-1}[\nabla\cdot(e^\theta v\nabla u_{f_\theta})]\|_{H^1}\\
&\lesssim \|\theta\|_\infty\|v\|_{C^1}\|\mathscr{L}_{f_\theta}^{-1}[\nabla\cdot(e^\theta v\nabla u_{f_\theta})]\|_{H^2}\\
&\lesssim \|v\|_{C^1}\|e^\theta v\nabla u_{f_\theta}\|_{H^1}\lesssim \|v\|_{C^1}\|v\|_{H^1}\|f_\theta\|_{C^1}\|u_{f_\theta}\|_{C^2}\\
&\lesssim \|v\|_{H^3}\|v\|_{H^1}\lesssim D^{4/d}\|v\|_{E_D}^2
\end{aligned}$$

where we have used the Sobolev imbedding $H^3\subset C^1$ for $d\le 3$ and again (5.51), (5.52), (6.15), (6.14).

Repeating arguments of this kind one can verify Condition 3.2.1 for Darcy's problem – the 'critical' term arises from the Lipschitz constant of the operator norms of $\nabla^2\mathscr{G}^x$, and the details of the proof of the following proposition are left to Ex. 5.4.4.

Proposition 5.3.4. *Under the hypotheses of Theorem 5.3.2, Condition 3.2.1 is satisfied for $\kappa_2 = 7/d$.*

No claim is made about the optimality of our bound for κ_2 – we only wish to exhibit its growth of a polynomial in D order to apply Theorem 5.1.3.

We have now everything at hand to apply Theorem 5.1.3 to Darcy's problem. The smoothness $\alpha>21$ of the model required in the following theorems is not optimal and arises as such because we have repeatedly opted for simpler proofs in the preceding results. Specifically we have not appealed to 'Schauder-theory' for C^α-regularity of solutions to elliptic PDEs (e.g., Section 6 in [60]) but only to Sobolev regularity bounds via L^2-energy estimates. When proofs require C^α-bounds, the use of the Sobolev imbedding then increases relevant constants. We did this to focus on the main ideas rather than on technicalities. The resulting assumption on α translates into a growth hypothesis on the dimension D, and the theorems that follow should hence be construed as holding in 'moderately high-dimensional' models – the precise value of α is not of central importance. Sharper versions are possible (and discussed in the notes) but do not impact the main insight drawn from the following theorem, that for sufficiently regular models, a log-concave approximation theorem is true for Darcy's problem while the Bernstein-von Mises theorem fails for it.

Theorem 5.3.5. *Suppose that $\theta_0\in\tilde H^\alpha(\mathcal{X})$ for $\alpha>21, d\le 3$, that (5.16) holds with $\delta_N = N^{-\alpha/(2\alpha+d)}$, and that (2.16) is satisfied for all $\theta\in\Theta\equiv\tilde H^\alpha$. Then the forward map $\mathscr{G}$ from (2.8) satisfies Condition 5.1.1 for $\mathbf{r}=rD^{-8/d}$ and all small enough $r>0$, spaces E_D from (5.3),*

$$\eta=\frac{\alpha-1}{\alpha+1},\ \kappa_0=\frac{6}{d},\ \kappa_1=0,\ \kappa_2=\frac{7}{d}$$

and some $c_i > 0, i = 0, 1, 2,$ *as long as* $D \leq A_1 N^{d/(2\alpha+d)}$ *for some* $A_1 < \infty$. *In particular the log-concave approximation from Theorem 5.1.3 holds true for the chosen constants and the corresponding surrogate posterior from (5.14) and with* $e^{-CND^{-2\kappa_0-4\kappa_2}} = o(1)$.

Proof. Note that Propositions 2.1.3 and 2.1.7 with $\beta = \alpha$ apply to verify Condition 2.1.1 for $\kappa = 1$ (and then also $\kappa = 0$) and Condition 2.1.4 with $\eta = (\alpha-1)/(\alpha+1)$. We take $\beta = \alpha$ here, permitted in view of Ex. 2.4.3 for this particular prior (5.4). Conditions 3.2.1 and 3.2.2 were checked in the previous subsection, using also that (5.16) verifies the bound on $\|\mathscr{G}(\theta_0) - \mathscr{G}(\theta_{0,D}\|_{L^2}$ for some c_1 and the choice of D to follow. Collecting the required bounds, we need $D^{-w} \gtrsim D^{(6+7)/d}\delta_N$ to verify (5.5) which is satisfied for our choice of D, α. $\square$

5.3.2 Polynomial time computation of the posterior mean

We can now combine the previous theorem with the computation time bounds from Section 5.2.1 for gradient based Langevin Markov chains. We only give a result for the polynomial time computatibility of the posterior mean vector – versions of Theorems 5.2.1 and 5.2.2 can be obtained as well. We consider a 'warm start' of the algorithm, assuming an appropriate initialiser θ_{init} into the region $\mathcal{B}_{\mathbf{r}}$ exists. When $D \approx N^a$ so that $\mathbf{r} \simeq N^{-\bar{a}}$ for some $a, \bar{a} > 0$, it is itself a non-trivial task to find a polynomial time initialiser – see the notes for discussion.

We recall that 'polynomial run-time' in the following theorem refers to the dimension D of the state space one is using, the 'informativeness' N of the posterior surface (5.22), as well as the desired precision level ε. Computational complexity is measured in terms of the number of required iterates of the MCMC scheme (5.23), including the burn-in time. Any MCMC step requires the evaluation of the gradient $\nabla \ell_N$ and hence of $\mathscr{G}, \nabla\mathscr{G}$, each amounting to the solution of an elliptic boundary value problem in $\mathcal{X} \subset \mathbb{R}^d$ (in the case of Darcy's problem). Such an operation is feasible (in polynomial time) by standard numerical PDE solvers, but we do not concern ourselves here with the computational cost of this operation in itself. The following theorem shows that computational barriers in D, N such as those described in Section 5.2.2 can be overcome for Darcy's problem by 'warm start' gradient type MCMC methods.

Theorem 5.3.6. *Suppose that data* $D_N = (Y_i, X_i)_{i=1}^N$ *arise from (1.11) in Darcy's problem with forward map* $\mathscr{G}$ *from (2.8),* $d \leq 3$, *and such that (2.16) is satisfied for all* $\theta \in \Theta \equiv \tilde{H}^\alpha, \alpha > 21$. *Consider the prior* Π *on* $E_D \subset \Theta$ *given by (5.4) with dimension* $D \leq A_1 N^{d/(2\alpha+d)}$ *for some* $A_1 > 0$. *Suppose that (5.16) holds for* $\theta_0 \in \Theta$. *Consider computation of the mean vector*

$$E^\Pi[\theta|D_N] = \int_{E_D} \theta d\Pi(\theta|D_N)d\theta$$

of the posterior distribution $\Pi(\cdot|D_N)$ *from (5.22). For some arbitrary fixed* $P > 0$ *let the precision level* ε *satisfy* $\varepsilon \geq N^{-P}$*. For an initialiser* θ_{init} *satisfying* $\|\theta_{\text{init}} - \theta_{0,D}\|_{E_D} \leq D^{-8/d}/(8\log N)$*, let the MCMC average* $\bar{\theta}^J_{J_{in}}$ *be as in Corollary 5.2.3 with* γ *as in (5.33) and corresponding choices of* m, Λ *in Theorem 5.2.1 provided by Theorem 5.3.5. Then there exist positive constants* a, b_1, b_2, b_3 *such that with* $P^N_{\theta_0} \times \mathbf{P}_{\theta_{\text{init}}}$*-probability at least* $1 - e^{N^a}$*,*

$$\left\|\bar{\theta}^J_{J_{in}} - E^\Pi[\theta|D_N]\right\|_{E_D} \leq \varepsilon,$$

and with burn-in and run-times J_{in}, J *of order at most polynomial order,*

$$J_{in} + J = O(N^{b_1} D^{b_2} \varepsilon^{-b_3}). \tag{5.55}$$

Proof. The result follows from Theorem 5.3.5 and Corollary 5.2.3 after bounding all relevant constants as discussed after (5.33). Note that the constants involving eigenvalues of the prior inverse covariance Σ^{-1} for our choice $\Sigma = diag(\lambda_j^\alpha : j \leq D)$ grow at most polynomially in D in view of (6.14). □

Note that the stepsizes γ in the previous algorithm are of the form $O(D^{-b_4} N^{-b_5} \epsilon^{b_6})$ for some $b_i > 0$.

We can combine the previous computational result with the consistency Theorem 2.3.3 for the posterior mean (or in fact Ex. 2.4.3) to further show that the MCMC averages also recover the ground truth θ_0 generating the data, that is, with $P^N_{\theta_0} \times \mathbf{P}_{\theta_{\text{init}}}$-probability at least $1 - e^{N^a}$ we have

$$\left\|\bar{\theta}^J_{J_{in}} - \theta_0\right\|_{L^2} \leq \varepsilon, \quad \text{whenever} \quad \varepsilon \gtrsim \max(N^{-P}, \delta_N^\eta), \tag{5.56}$$

after number $J + J_{in}$ of iterates of the MCMC chain that grows at most polynomially in relevant quantites (cf. (5.55)). This shows that the ground truth diffusivity θ_0 in Darcy's problem can be recovered by a constructive polynomial time MCMC algorithm, assuming warm start initialisation as in (5.6) is possible.

5.4 Notes

5.4.1 Exercises

Exercise 5.4.1. Prove Proposition 5.1.2. [Hint: A proof of a more general result can be found in Proposition 3.6 in [102]]

Exercise 5.4.2. Show that the probability bound in Lemma 1.3.3 can be taken of order $1 - O(e^{-bN\delta_N^2})$ for some $b > 0$ if K is chosen large enough. [Hint: Use (6.48) or see Lemma 4.15 in [102].] Further show that (5.20) holds.

Exercise 5.4.3. Complete the proof of Proposition 5.3.1 and show further that the map $\theta \mapsto \mathscr{G}^x(\theta)$ is C^3 on the set $\mathcal{B} = \mathcal{B}_{\mathbf{r}}$ from Theorem 5.3.2. [See also [4].]

Exercise 5.4.4. Prove Proposition 5.3.4. [See also [4].]

Exercise 5.4.5. Prove an analogue of Theorem 5.3.2 for the forward map $\mathscr{G}$ from the Schrödinger model (2.5), with $w = 4/d$ and lower bound in (5.50) of oder $c_2 D^{-4/d}$. [Hint: Start from (3.60), or see Lemma 4.7 in [102].]

5.4.2 Remarks and comments

The main ideas of this section were developed in the recent contribution [102] and in the follow up paper [23], even though the particular proofs given here follow a slightly different route inspired by the very recent article [4] which avoids the analysis of 'MAP' estimators from [23, 102] at the expense of slightly stronger conditions on the model parameters in Condition 5.1.1. The mixing time results from Subsection 5.2.1 follow ideas from [39, 49, 50] for strongly log-concave targets (reviewed in the appendix).

The results in [102] and [23] consider different PDEs than Darcy's problem, specifically the Schrödinger equation from (2.5) and non-Abelian X-ray transforms from (1.1), where results similar to Theorems 5.3.5 and 5.3.6 are obtained. In particular these references establish versions of the local average curvature Theorem 5.3.2 for these PDEs (with appropriate choices of $\kappa_0, \mathbf{r}, E_D$). The underlying gradient stability requirement seems more feasible to check than solving information equations (3.16), as highlighted by Darcy's problem. We also note that the choice of the discretisation spaces E_D needs to be adapted to the mapping properties of $D\mathscr{G}_\theta$ and for instance for X-ray transforms, an eigen-basis different from the standard Laplacian has to be considered to adjust for boundary behaviour, see [23], building on ideas from [89, 92].

As mentioned above, the proofs of the analogues of Theorem 5.1.3 in [23, 102] are slightly different in that they require a separate convergence analysis of the MAP estimator $\hat{\theta}_{MAP}$ (i.e., the maximiser of the posterior density of (5.22) over E_D). The benefit of this is that when dealing with term II in the decomposition of the Wasserstein distance in the proof of Theorem 5.1.3, one can expand the log-posterior around the global maximum $\hat{\theta}_{MAP}$ (instead of $\theta_{0,D}$), avoiding the gradient term in (5.19). This allows to weaken the quantitative requirements in Condition 5.1.1 – specifically the factor $D^{\kappa_0+\kappa_2}$ in the definition of $\tilde{\delta}_N$ can be replaced by just $D^{\kappa_0/2}$, see Condition 3.5 in [23] – but at the cost of introducing slightly stronger analytical conditions on $\mathscr{G}$ required to obtain convergence rates for $\hat{\theta}_{MAP}$. In this context [102] also provide computational guarantees for optimisation based estimators and further clarify that the unimodal 'surrogate posterior' actually has

$\hat{\theta}_{MAP}$ as its mode, as one may expect. While these stronger sets of conditions on $\mathscr{G}$ can often be checked for concrete PDEs, the proofs presented in the notes here are 'entirely Bayesian' and do not require to set up the machinery from M-estimation [125], [101] to study optimisation based estimators.

Log-concave approximation theorems such as Theorem 5.1.3 provide a conceptual alternative to Bernstein-von Mises approximations studied in Section 4.1 and also to Laplace approximations discussed in Remark 5.1.4. They are useful to establish computational guarantees for gradient based MCMC methods as in Theorem 5.3.6 – a main challenge here is to show that the MCMC method requires at most a *polynomial* run-time *both in D and N simultaneously.* Results that scale uniformly in D (but exponentially in N) can be obtained by infinite-dimensional Harris-type theorems – see [66] - but these do not give polynomial time guarantees for 'informative' posteriors when $N \to \infty$. See [102] for more discussion. We note also that log-concave approximation theorems *appear to fall short* (at least in our current understanding) of providing *statistical* guarantees for posterior-based uncertainty quantification such as those given in Section 4.1.3. The reason for this is a) that Theorem 5.1.3 is not compatible with the weak convergence arguments relevant in Section 4.1.3 (as no N-independent limiting measure is provided), and b) that as such the Wasserstein-distance does not dominate the Kolmogorov 'quantile' distance for general pairs of N-dependent probability measures. Thus for purposes of uncertainty quantification, the BvM-theorems from Section 4.1 are currently still the only available results, despite their limitations.

The results in Subsection 5.2.2 about hardness of local MCMC methods are partly inspired by ideas from [15, 16] (see also [10]) laid out in very different high-dimensional statistical physics models. Theorem 5.2.4 presented here is a simplified 'average log-likelihood' versions of one of the main results from [11], where also other priors and random step size MCMC methods such as MALA and pCN are explicitly considered.

We finally touch on the issue of initialisation in (5.6), specifically with an eye on data-driven choices for θ_{init} as is relevant in Theorem 5.3.6 on the performance of gradient MCMC. We are not aware of a general strategy for initialisation other than perhaps in settings where the curvature hypothesis in Condition 3.2.2 holds on a fixed ball $\mathcal{B}$ (of constant in D, N radius) when a standard grid-search can be used. For the Schrödinger equation, an ad hoc strategy for initialisation can be provided [102], but for other PDEs this poses novel challenges that are not sufficiently well understood at the time of this writing.

Chapter 6

Appendix

6.1 Analytical background

In this section, unless mentioned otherwise, $\mathcal{X}$ is always a bounded domain in $\mathbb{R}^d$ with smooth boundary and the space $L^2 = L^2(\mathcal{X}) = L^2_\lambda(\mathcal{X})$ denotes the standard Hilbert space of square integrable functions on $\mathcal{X}$ with respect to Lebesgue measure λ, with inner product $\langle \cdot, \cdot \rangle_{L^2}$ and norm $\| \cdot \|_{L^2} = \sqrt{\langle \cdot, \cdot \rangle_{L^2}}$. The relation $a \lesssim b$ denotes an inequality $a \leq Cb$ that holds up to fixed constants $C > 0$, and the corresponding convention is used for $\gtrsim$.

6.1.1 Sobolev and related spaces

We collect here without proofs some standard facts about Sobolev spaces that can be found in any of the standard references on the subject, for instance Chapter 1 in [87], Chapter 4 in [122], Chapter 7 in [60] or [2].

For a multi-index $i = (i_1, ..., i_d)$, let D^i denote the i-th (weak) partial derivative operator of order $|i|$. Then for integer $\alpha \geq 0$, the Sobolev spaces are

$$H^\alpha(\mathcal{X}) := \left\{ f \in L^2(\mathcal{X}) \;\middle|\; \text{for all } |i| \leq \alpha,\ D^i f \text{ exists and } D^i f \in L^2(\mathcal{X}) \right\}, \tag{6.1}$$

normed by

$$\|f\|^2_{H^\alpha(\mathcal{X})} = \sum_{|i| \leq \alpha} \|D^i f\|^2_{L^2(\mathcal{X})}.$$

For non-integer real values $\alpha \geq 0$ one can define $H^\alpha(\mathcal{X})$ by interpolation. For $\alpha = 0$ the convention $H^0(\mathcal{X}) = L^2(\mathcal{X})$ will be used.

The space of bounded and continuous functions on $\mathcal{X}$ is denoted by $C(\mathcal{X})$, equipped with the supremum norm $\| \cdot \|_\infty$. For $\eta \in \mathbb{N}$, the space of η-times differentiable functions on $\mathcal{X}$ with (bounded) uniformly continuous derivatives is denoted by $C^\eta(\mathcal{X})$, with norm $\|f\|_{C^\eta(\mathcal{X})} = \sum_{|i| \leq \eta} \|D^i f\|_\infty$. For $\eta > 0, \eta \notin \mathbb{N}$, we say $f \in C^\eta(\mathcal{X})$ if for all multi-indices β with $|\beta| \le \lfloor \eta \rfloor$ (the integer part of η), $D^\beta f$ exists and is $\eta - \lfloor \eta \rfloor$-Hölder continuous. The norm on $C^\eta(\mathcal{X})$ for such η is

$$\|f\|_{C^\eta(\mathcal{X})} = \sum_{\beta:|\beta|\le\lfloor\eta\rfloor} \|D^\beta f\|_\infty + \sum_{\beta:|\beta|=\lfloor\eta\rfloor} \sup_{x,y\in\mathcal{X},\, x\neq y} \frac{|D^\beta f(x) - D^\beta f(y)|}{|x-y|^{\eta-\lfloor\eta\rfloor}}.$$

We also define the set of smooth functions as $C^\infty(\mathcal{X}) = \cap_{\eta>0} C^\eta(\mathcal{X})$ and its subspace $C_c^\infty(\mathcal{X})$ of functions compactly supported in $\mathcal{X}$. All definitions so far make sense for any open subset $\mathcal{X} \subseteq \mathbb{R}^d$.

Now for $\mathcal{X} \subset \mathbb{R}^d$ a bounded domain with smooth boundary $\partial\mathcal{X}$, the completion of $C_c^\infty(\mathcal{X})$ for the H^α-norm is denoted by $H_c^\alpha(\mathcal{X})$, and this is a closed subspace of $H^\alpha(\mathcal{X})$ (equal to $H^\alpha(\mathcal{X})$ only if $\alpha < 1/2$). We also define the spaces $H_0^\alpha(\mathcal{X}), C_0^\infty(\mathcal{X})$ of elements of $H^\alpha(\mathcal{X}), C^\infty(\mathcal{X})$ that vanish at $\partial\mathcal{X}$ (in a trace sense if necessary). We notice that $H_0^1(\mathcal{X}) = H_c^1(\mathcal{X})$ but otherwise ($\alpha > 1$) the inclusion $H_c^\alpha(\mathcal{X}) \subset H_0^\alpha(\mathcal{X})$ is strict as the former spaces require normal derivatives up to order $\alpha - 1/2$ to vanish at $\partial\mathcal{X}$. One also defines topological dual spaces $H^{-\alpha}(\mathcal{X}) = (H_c^\alpha(\mathcal{X}))^*$ for $\alpha \ge 0$. For some results we also need Sobolev spaces $H^\alpha(\partial\mathcal{X})$ defined similarly on the closed 'boundary-less' manifold $\partial\mathcal{X}$, as well as $C^\infty(\partial\mathcal{X}) = \cap_{\alpha\ge0} H^\alpha(\partial\mathcal{X})$. See, e.g., Chapter 4.3 in [122] for details.

All preceding spaces and norms can be defined for vector fields $f : \mathcal{X} \to \mathbb{R}^d$ with standard modification of the norms, by requiring each of the coordinate functions $f_i(\cdot), i = 1, \ldots, d$, to belong to the corresponding space of real-valued maps.

We will repeatedly use the standard multiplicative inequalities

$$\|fg\|_{H^\alpha} \lesssim \|f\|_{H^\alpha}\|g\|_{H^\alpha}, \qquad \alpha > d/2, \tag{6.2}$$
$$\|fg\|_{H^\alpha} \lesssim \|f\|_{C^\alpha}\|g\|_{H^\alpha}, \qquad \alpha \ge 0, \tag{6.3}$$

for all f, g in the appropriate function spaces. Also, for any $\alpha > d/2$ and $0 < \eta < \alpha - d/2$, the Sobolev embedding $H^\alpha \subseteq C^\eta$ holds, with corresponding norm estimates

$$\forall f \in H^\alpha, \ \ \|f\|_\infty \lesssim \|f\|_{C^\eta} \lesssim \|f\|_{H^\alpha}. \tag{6.4}$$

We also need the following interpolation inequality for Sobolev norms. For all $\beta_1, \beta_2 \ge 0$ and $\theta \in [0, 1]$,

$$\forall u \in H^{\beta_1} \cap H^{\beta_2} : \quad \|u\|_{H^{\theta\beta_1+(1-\theta)\beta_2}} \lesssim \|u\|_{H^{\beta_1}}^\theta \|u\|_{H^{\beta_2}}^{1-\theta}. \tag{6.5}$$

6.1.2 Elliptic second order differential operators

We now consider a second order elliptic operator

$$L_{\gamma,V}(u) = \nabla \cdot (\gamma \nabla u) - Vu, \quad u \in C^\infty(\mathcal{X}),$$

where $\gamma : \mathcal{X} \to [\gamma_{min}, \infty)$ is a diffusion coefficient bounded from below by a scalar $\gamma_{min} > 0$, $V : \mathcal{X} \to [0, \infty)$ a non-negative potential, and $\nabla \cdot v = \sum_{j=1}^d \partial v/\partial v_j$ denotes the divergence of a vector field $v \in H^1(\mathcal{X})$. We will assume throughout this section that $\gamma \in H^\beta(\mathcal{X})$ for some $\beta > 1 + d/2$ so that in particular $\gamma \in C^\eta(\mathcal{X})$ for $\eta > 1$ by (6.4). We also assume that $V \in H^{\beta-1}(\mathcal{X}) \subset C^{\eta-1}(\mathcal{X})$. Weaker regularity hypotheses on γ, V are possible but we do not pursue those to facilitate the exposition.

With this setup we can represent the divergence form operator $\mathscr{L}_f$ from (1.3) as $L_{f,0}$ and the Schrödinger operator $\mathscr{L}_f$ from (1.4) as $L_{1/2,f}$. The PDE (1.2) is thus a special case of the more general PDE

$$\begin{aligned} L_{\gamma,V} v &= g \quad \text{on } \mathcal{X}, \\ v &= h \quad \text{on } \partial\mathcal{X}, \end{aligned} \tag{6.6}$$

whose solutions v we now study. We assume again that g, h are given smooth (C^∞-) functions on $\mathcal{X}, \partial\mathcal{X}$, respectively.

For a smooth vector field $v : \mathcal{X} \to \mathbb{R}^d$ the divergence theorem (Prop. 2.3 on p.143 in [122] or p.714f. in [53]) implies that for any $u \in C_0^\infty(\mathcal{X})$

$$\int_{\mathcal{X}} u \nabla \cdot v d\lambda = -\int_{\mathcal{X}} \nabla u \cdot v d\lambda \tag{6.7}$$

and this extends to $v \in H^1(\mathcal{X}), u \in H_0^1(\mathcal{X})$ by approximation.

Now using (6.7) and the Poincaré inequality (e.g., p. 340 in [122]), for any $u \in H_0^1(\mathcal{X})$

$$-\langle L_{\gamma,V} u, u \rangle_{L^2(\mathcal{X})} = \langle \gamma \nabla u, \nabla u \rangle_{L^2} + \langle V u, u \rangle_{L^2} \geq \gamma_{min} \|\nabla u\|_{L^2}^2 \geq C(\gamma_{min}, \mathcal{X}) \|u\|_{H^1}^2.$$

Using the dual characterisation of the norm of $H^{-1}(\mathcal{X}) = (H_0^1(\mathcal{X}))^*$ and testing with $\phi = -u$ we deduce

$$\|L_{\gamma,V} u\|_{H^{-1}} = \sup_{\|\phi\|_{H_0^1} \leq 1} \langle L_{\gamma,V} u, \phi \rangle_{L^2} \geq C(\gamma_{min}, \mathcal{X}) \|u\|_{H^1}. \tag{6.8}$$

Similar arguments imply a converse inequality of (6.8) and hence that $L_{\gamma,V}$ continuously maps $H_0^1(\mathcal{X})$ into $H^{-1}(\mathcal{X})$ with operator norm bounded by some $D = D(\|\gamma\|_\infty, \|V\|_\infty)$. From this one then deduces the following facts, see e.g., p.355f. in [122]: the operator $L_{\gamma,V}$ realises an isomorphism of $H_0^1(\mathcal{X})$ onto $H^{-1}(\mathcal{X})$ with a well defined inverse $L_{\gamma,V}^{-1}$ of $L_{\gamma,V}$ that defines a compact negative self-adjoint operator on the Hilbert space $L_\lambda^2(\mathcal{X})$, mapping into $H_0^1(\mathcal{X})$. Compactness and the spectral theorem furnish us with eigen-pairs

$$(\lambda_j, e_j) = (\lambda_{\gamma,V,j}, e_{j,\gamma,V}) \in (0, \infty) \times H_0^1(\mathcal{X}), j \in \mathbb{N}, \tag{6.9}$$

of $-L_{\gamma,V}$ that depend on γ, V (suppressed in the notation unless necessary). In particular the action of $L_{\gamma,V}$ on sufficiently regular elements of $L^2(\mathcal{X})$ can be represented as in (1.6), and for $h = 0$ and $g \in L^2(\mathcal{X})$ we can represent the unique solutions $v \in H_0^1(\mathcal{X})$ of (6.6) as

$$v = L_{\gamma,V}^{-1}[g] = -\sum_j \lambda_j^{-1} e_j \langle e_j, g\rangle_{L^2}. \tag{6.10}$$

Unique solutions v in $H^1(\mathcal{X})$ for (6.6) with $h \neq 0$ can be constructed likewise (see Chapter 8 in [60] or arguing as on p.359f. in [122]), and for smooth g, h such v will belong to a Sobolev space if γ, V do: Theorems 8.12 and 8.13 in [60] imply that

$$v \in H^{k+1}(\mathcal{X}) \text{ if } \gamma \in H^k(\mathcal{X}), V \in H^{k-1}(\mathcal{X}), \quad k \in \mathbb{N}, \tag{6.11}$$

and that this remains true as long as the source g belongs to H^{k-1} (rather than being smooth).

We will show below (Proposition 6.1.6) that $L_{\gamma,V}^{-1}$ is in fact a Lipschitz operator between such Sobolev spaces but need to be careful to track the constants in dependence on V, γ, k. Some first Lipschitz estimates that are uniform in V, γ already follow from (6.8) with $u = L_{\gamma,V}^{-1}[g]$, specifically since

$$\|u\|_{L^2} \leq \|u\|_{H^1} \leq c\|L_{\gamma,V} u\|_{H^{-1}} \leq c\|L_{\gamma,V} u\|_{L^2}, \quad u \in H_0^1,$$

for some $c = c(\gamma_{min}, \mathcal{X}) < \infty$, we see that for any $g \in L^2$,

$$\|L_{\gamma,V}^{-1}[g]\|_{L^2} \leq c\|g\|_{L^2}, \tag{6.12}$$

and we also have the stronger estimate

$$\|L_{\gamma,V}^{-1}[g]\|_{H^1} \leq c\|g\|_{L^2}. \tag{6.13}$$

6.1.3 Orthonormal discretisation of L^2 and metric entropy

The spectrally defined Sobolev-type spaces $\tilde{H}^s$

The spectral theorem furnishing the representation (6.10) in fact implies that $\{e_j : j \in \mathbb{N}\}$ is a complete $\langle \cdot, \cdot \rangle_{L^2}$-orthonormal basis of the Hilbert space $L^2 = L^2_\lambda(\mathcal{X})$. Setting $\gamma = 1, V = 0$, the well-known Weyl-asymptotics for the eigenfunctions of the standard Laplacian $-\Delta = -L_{1,0}$ are

$$\lambda_j \simeq j^{2/d}, \quad j \to \infty, \tag{6.14}$$

see [123], Corollary 8.3.5. For these eigen-pairs of Δ, new Sobolev-type spaces are then defined as

$$\tilde{H}^s(\mathcal{X}) = \Big\{ f : \|f\|^2_{\tilde{H}^s(\mathcal{X})} := \sum_{j \in \mathbb{N}} \lambda_j^s \langle f, e_j\rangle^2_{L^2} < \infty \Big\}, \; s \in \mathbb{R}, \tag{6.15}$$

which carry a natural (Hilbert-space) inner product

$$\langle f, g\rangle_{\tilde{H}^s} = \sum_j \lambda_j^s \langle f, e_j\rangle_{L^2}\langle g, e_j\rangle_{L^2}.$$

For $s \geq 0$ the above spaces consist of elements of $L^2(\mathcal{X})$ but for $s < 0$ these are generalised functions (in the Schwartz sense) and a basic sequence space duality argument shows that

$$\tilde{H}^s(\mathcal{X}) = (\tilde{H}^{-s}(\mathcal{X}))^*, \quad s < 0, \tag{6.16}$$

we refer to p.473 in [122] for details.

One can define the $\tilde{H}^s(\gamma)$ spaces also for the spectrum of the $L_{\gamma,0}$-operators instead of the standard Laplacian $\gamma = 1$ – the sequence norms of these $\tilde{H}^s(\gamma), s = 0, 1, 2,$ spaces can all be shown to be equivalent with equivalence constants depending only on $U \geq \|\gamma\|_{C^1}$ and $\gamma_{min} \leq \gamma \leq \gamma_{max}$. Further properties of these spaces are discussed after the next subsection.

The metric entropy bound in sequence space

A lemma used repeatedly in these notes is the following combinatorial result for the metric complexity of balls in $\tilde{H}^\alpha$ spaces.

For a totally bounded subset A of a metric space (X, d_X), the number $N(A, d_X, \epsilon)$ denotes the minimal number of d_X-balls of radius $\epsilon > 0$ required to cover A, and $\log N(A, d_X, \epsilon)$ is called the *metric entropy* of A. When d_X arises from a norm $\|\cdot\|$ we just write $N(A, \|\cdot\|, \epsilon)$ instead.

Proposition 6.1.1. *Let $\alpha \geq 0$ and $-\infty < k < \alpha$. Then the ϵ-log-covering numbers of the ball $\tilde{h}^\alpha(B)$ of $\tilde{H}^\alpha(\mathcal{X})$-norm radius B for the $\tilde{H}^k(\mathcal{X})$-distance satisfy*

$$\log N\big(\tilde{h}^\alpha(B), \|\cdot\|_{\tilde{H}^k}, \epsilon\big) \leq \Big(\frac{AB}{\epsilon}\Big)^{d/(\alpha-k)}, \quad 0 < \epsilon < AB, \tag{6.17}$$

where $A = A(d, \alpha, k) < \infty$ is a fixed constant.

Proof. We first prove the case $k = 0$. A standard scaling argument for norms allows to restrict to $B = 1$. Let us write $f_n = \langle f, e_n\rangle_{L^2}$ for the 'Fourier' coefficients of $f \in L^2(\mathcal{X})$ in the basis e_n throughout the proof. We also note that (6.14) provides a constant $b_1 > 0$ such that $b_1 j^{2/d} \leq \lambda_j$ for all $j \in \mathbb{N}$.

For any f contained in

$$\tilde{h}^\alpha(1) = \Big\{f : \sum_n \lambda_n^\alpha f_n^2 \leq 1\Big\},$$

using (6.14) and Parseval's identity, we can estimate the error of L^2_ζ-approximation $P_{2^{\ell_0}} f$ of f from all frequencies up to the largest integer $n < 2^{\ell_0}$, as

$$\|f - P_{2^{\ell_0}} f\|^2_{L^2_\zeta} = \sum_{n \geq 2^{\ell_0}} f_n^2 \lambda_n^\alpha \lambda_n^{-\alpha} \leq b_1^{-\alpha} 2^{-2\alpha\ell_0/d} \leq (\epsilon/4)^2 \tag{6.18}$$

where we have chosen ℓ_0 as

$$\ell_0 = \frac{d}{\alpha} \log_2 \Big(\frac{4 \cdot 2^{(\alpha/d)+1} b_1^{-\alpha/2}}{\epsilon} \Big) \tag{6.19}$$

The remaining indices $n : 0 \leq n < 2^{\ell_0}$ are now decomposed into dyadic brackets

$$N_l = \{n : 2^{l-1} \leq n \leq 2^l - 1\}, \ l = 1, \dots, ; |N_l| = 2^{l-1},$$

where by convention we set $N_0 = \{0\}$. Thus using again Parseval's identity, and the preceding estimate we can bound, for any $f, g \in \tilde{h}^\alpha(1)$,

$$\begin{aligned} \|f - g\|_{L^2_\zeta} &\leq \sqrt{\sum_{0 \leq l \leq \ell_0 + 1} \sum_{n \in N_l} \lambda_n^{-\alpha} \lambda_n^\alpha (f_n - g_n)^2} + 2\epsilon/4 \\ &\leq 2^{\alpha/d} b_1^{-\alpha/2} \sum_{0 \leq l \leq \ell_0 + 1} 2^{-l\alpha/d} \sqrt{\sum_{n \in N_l} (\lambda_n^{\alpha/2} f_n - \lambda_n^{\alpha/2} g_n)^2} + \epsilon/2. \end{aligned}$$

Now by definition of $\tilde{h}^\alpha(1)$, for every $l \in \mathbb{N} \cup \{0\}$ the vectors $\{\lambda_n^{\alpha/2} f_n : n \in N_l\}$, $\{\lambda_n^{\alpha/2} g_n : n \in N_l\}$ lie in the unit ball of a Euclidean space of dimension $|N_l| \leq 2^l$. The Euclidean ϵ'-covering numbers of such a unit ball B_l for the standard Euclidean norm are

$$N(l) \equiv N(B_l, \|\cdot\|_{\mathbb{R}^{|N_l|}}, \epsilon') \leq (3/\epsilon')^{2^l}, \ 0 < \epsilon' < 1,$$

(see, e.g., Proposition 4.3.34 in [61]). By choosing radius

$$\epsilon'_l = 2^{l((\alpha/d)+1)} 2^{-\ell_0((\alpha/d)+1)}$$

coverings of B_l for each l centred at points $(f_{n,l,i} : n \in N_l, l \leq \ell_0 + 1)$, and setting $f_{n,l,i} = 0$ for $n \geq 2^{\ell_0+1}$, we obtain a covering

$$\Big(\bar{f}^{(i)} = \lambda_n^{-\alpha/2} f_{n,l,i} : i = 1, \dots, \prod_{0 \leq l \leq \ell_0+1} N(l) \Big)$$

of $\tilde{h}^\alpha(1)$ of radius bounded by

$$\begin{aligned} \|f - \bar{f}^{(i)}\|_{L^2_\zeta} &\leq \frac{\epsilon}{2} + 2^{\alpha/d} b_1^{-\alpha/2} \sum_{0 \leq l \leq \ell_0 + 1} 2^{-l\alpha/d} 2^{l((\alpha/d)+1)} 2^{-\ell_0((\alpha/d)+1)} \\ &\leq \frac{\epsilon}{2} + 2^{(\alpha/d)+2} b_1^{-\alpha/2} 2^{-\alpha\ell_0/d} \leq \epsilon, \end{aligned}$$

by choice of ℓ_0. We thus obtain, for some $c' = c'(\alpha, d) > 0$,

$$\log_2 N(\tilde h^\alpha(1), \|\cdot\|_{L^2_\zeta}, \epsilon) \le \sum_{0\le l\le \ell_0+1} \log_2 N(l) \le \sum_{0\le l\le \ell_0+1} 2^l \Big[\log_2 3 + (\ell_0 - l)\big(\frac{\alpha}{d}+1\big)\Big]$$
$$\le c' 2^{\ell_0} \le (A/\epsilon)^{d/\alpha}$$

so that (6.17) follows for $k=0$.

For $-\infty < k < \alpha$, if $f, g \in \tilde h^\alpha(B)$ we can write

$$\|f-g\|^2_{\tilde H^k} = \sum_n (f_n \lambda_n^{k/2} - g_n \lambda_n^{k/2})^2 = \|\tilde f - \tilde g\|^2_{L^2} \tag{6.20}$$

where $\tilde f, \tilde g$ are defined via 'Fourier coefficients' $\lambda_n^{k/2} f_n, \lambda_n^{k/2} g_n, n \in \mathbb{N}$ which belong to a bounded subset of $\tilde H^{\alpha-k}$. Hence a $\|\cdot\|_{L^2}$-covering of $\{\tilde f : f \in \tilde h^\alpha(B)\} \subset \tilde h^{\alpha-k}(B)$ induces a $\tilde H^k$-covering of $h^\alpha(B)$, and the bound for $k=0$ just established applies with $\alpha - k$ in place of α, completing the proof. □

Remark 6.1.2. One can show that the inequality (6.17) is optimal in the sense that a corresponding lower bound holds with A replaced by some $A' > 0$ depending only on d, α, k, see Chapter 3 in [51] for a proof.

Metric entropy estimates for standard Sobolev spaces

The spaces $\tilde H^\alpha(\mathcal{X})$ have to be understood relative to the 'eigen-pairs' (e_n, λ_n) of the Dirichlet Laplacian $-\Delta = L_{1,0}$ on $\mathcal{X}$ and (unlike on boundary-less manifolds) are *not* equal to the standard spaces $H^\alpha(\mathcal{X})$, due to the presence of a boundary. However, arguing as in Section 5.A in [122], one shows the following facts: We have

$$\tilde H^\alpha(\mathcal{X}) = H^\alpha_0(\mathcal{X}), \alpha = 1, 2; \quad \tilde H^\alpha(\mathcal{X}) \subset H^\alpha_0(\mathcal{X}), \alpha \in \mathbb{N}, \tag{6.21}$$

and the $\|\cdot\|_{H^\alpha}$ and $\|\cdot\|_{\tilde H^\alpha}$-norms are Lipschitz equivalent on $\tilde H^\alpha$. Moreover any $f \in H^\alpha(\mathcal{X})$ that is supported in some compact subset $K \subset \mathcal{X}$ belongs to $\tilde H^\alpha(\mathcal{X})$ with norm estimate $\|f\|_{\tilde H^\alpha} \le c_K \|f\|_{H^\alpha}$. This is clear for $\alpha = 1, 2$ and extends to $\alpha \in \mathbb{N}$ by induction as $\tilde H^\alpha(\mathcal{X})$ is the image of $H^{\alpha-2}(\mathcal{X})$ under an application of $L^{-1}_{1,0}$, cf. (6.32). It extends in fact also to $\alpha < 0$ in that any element of $H^\alpha(\mathcal{X}) := (H^{-\alpha}_c(\mathcal{X}))^*$ that is compactly supported in $\mathcal{X}$ in the sense of Schwartz distributions belongs to $\tilde H^\alpha(\mathcal{X})$, see p.474 in [122].

From what precedes and Proposition 6.1.1 one can obtain the classical L^2-metric entropy inequality for balls $h^\alpha(B)$ of radius B in $H^\alpha(\mathcal{X})$

$$\log N(h^\alpha(B), \|\cdot\|_{L^2(\mathcal{X})}, \epsilon) \lesssim \Big(\frac{B}{\epsilon}\Big)^{d/\alpha}, \ 0 < \epsilon < B, \tag{6.22}$$

by an extension argument: Let $\mathcal{Y} \supset \mathcal{X}$ be a bounded smooth domain in $\mathbb{R}^d$ such that that the closure of $\mathcal{X}$ is contained in the interior of $\mathcal{Y}$. Elements f of $H^\alpha(\mathcal{X})$ can be regarded as restrictions of functions $\bar{f}: \mathcal{Y} \to \mathbb{R}$ in $H^\alpha(\mathcal{Y})$ that are compactly supported in $\mathcal{Y}$ and for which

$$\|\bar{f}\|_{\tilde{H}^\alpha(\mathcal{Y})} \lesssim \|\bar{f}\|_{H^\alpha(\mathcal{Y})} \lesssim \|f\|_{H^\alpha(\mathcal{X})};$$

the second inequality follows from the usual extension theorem for Sobolev functions (see Chapter 1.8 in [87]) and by employing a 'cut-off' function in $C_c^\infty(\mathcal{Y})$ that equals one identically on $\mathcal{X}$. Thus an ϵ-$L^2(\mathcal{Y})$-covering of a sufficiently large ball in $\tilde{H}^\alpha(\mathcal{Y})$ induces a covering of a ball in $H^\alpha(\mathcal{X})$ of $L^2(\mathcal{X})$ radius

$$\|f - g\|_{L^2(\mathcal{X})} \le \|\bar{f} - \bar{g}\|_{L^2(\mathcal{Y})} \le \epsilon, \tag{6.23}$$

and (6.22) then follows from Proposition 6.1.1.

When covering a ball in $H_c^\alpha(\mathcal{X})$, the previous extension argument can be refined to cover a weaker class of norms (relevant in Condition 2.1.1). Elements $f \in H_c^\alpha(\mathcal{X})$ vanish at $\partial\mathcal{X}$ in a sufficiently regular way that their extension $\tilde{f}$ *by zero* outside of $\mathcal{X}$ belongs to $H^\alpha(\mathcal{Y})$ (see [87], Theorem 1.11.4). If we consider the topological dual space $(\tilde{H}^\kappa(\mathcal{X}))^*$ for $\kappa \ge 0$, then

$$\|f\|_{(H^\kappa(\mathcal{X}))^*} = \sup_{\|\phi\|_{H^\kappa(\mathcal{X})}\le 1} \left|\int_{\mathcal{X}} \phi f\right| \le \sup_{\bar{\phi}: \|\bar{\phi}\|_{H^\kappa(\mathcal{Y})}\le C} \left|\int_{\mathcal{Y}} \bar{\phi}\tilde{f}\right| \lesssim \|\tilde{f}\|_{\tilde{H}^{-\kappa}(\mathcal{Y})} \tag{6.24}$$

where the last supremum ranges over those $\bar{\phi}$ which satisfy $\bar{\phi} = \phi$ on $\mathcal{X}$ and are compactly supported in $\mathcal{Y}$, whence $\|\bar{\phi}\|_{\tilde{H}^\kappa(\mathcal{Y})} \le C'$ for some $C' > 0$, and where the last inequality follows from (6.16). We therefore deduce from Proposition 6.1.1 that a ball $h_c^\alpha(B)$ of radius B in $H_c^\alpha(\mathcal{X})$ satisfies the metric entropy inequality

$$\log N(h_c^\alpha(B), \|\cdot\|_{(H^\kappa(\mathcal{X}))^*}, \epsilon) \lesssim \left(\frac{B}{\epsilon}\right)^{d/(\alpha+\kappa)}, \; 0 < \varepsilon < B. \tag{6.25}$$

For $0 \le \kappa < 1/2$ the last inequality holds for $h^\alpha(B) \supset h_c^\alpha(B)$ – the boundary constraint can be relaxed as we can then use $\bar{f}$ instead of $\tilde{f}$ in (6.24) together with the fact that zero extensions $\tilde{\phi}$ of $\phi \in H^\kappa(\mathcal{X})$ belong to $H^\kappa(\mathcal{Y})$ and then also $\tilde{H}^\kappa(\mathcal{Y})$ for such κ (see [87], Theorem 1.11.4).

Results of the type given in this subsection for general function spaces can be found in [51] but we prefer to give self-contained proofs of the precise bounds required in our setting.

6.1.4 Feynman-Kac formulæ

Solutions v of elliptic PDEs (6.6) have probabilistic representations by virtue of integrals against sample paths of diffusion processes, e.g., [13, 56]. For the applications of diffusion process techniques it is convenient to ensure that C^2-solutions

v of (6.6) exist. In our setting where $(\gamma, V) \in H^\beta \times H^{\beta-1}$ for some $\beta > 1+d/2$, the regularity result (6.11) with $k = \beta$ and the Sobolev imbedding (6.4) are sufficient to ensure $v \in H^{\beta+1} \subset C^2$.

From the theory of stochastic differential equations (e.g., p.10 in [13], and noting $\gamma \in H^\beta \subset C^1$ so that also $\sqrt{\gamma} \in C^1$ as $\gamma \geq \gamma_{min} > 0$) there exists a unique Markov diffusion process $(X_t : t \geq 0)$ on $\mathbb{R}^d$ which has $\mathscr{L}_{\gamma,0} = \nabla \cdot (\gamma \nabla)$ as a generator, corresponding to path-wise solutions to the stochastic differential equation

$$dX_t = \nabla\gamma(X_t)dt + \sqrt{2\gamma(X_t)}dW_t, \quad t \geq 0, \tag{6.26}$$

where $(W_t : t \geq 0)$ is a d-dimensional Brownian motion. Then by Theorem 2.1 on p.127 in [56], the C^2-solution v to the elliptic PDE (6.6) has representation

$$v(x) = -E_\gamma^x\Big[\int_0^{\tau_{\mathcal{X}}} g(X_t)e^{-\int_0^t V(X_s)ds}dt\Big] + E_\gamma^x\Big[h(X_{\tau_{\mathcal{X}}})e^{-\int_0^{\tau_{\mathcal{X}}} V(X_s)ds}\Big], \tag{6.27}$$

for every $x \in \mathcal{X}$, where E_γ^x is the expectation under the law of the process $(X_t : t \geq 0)$ solving (6.26) started at x with exit time $\tau_{\mathcal{X}}$ from the domain $\mathcal{X}$.

Lemma 6.1.3. *We have* $\sup_{x\in\mathcal{X}} E_\gamma^x \tau_{\mathcal{X}} \leq c$ *for some* $c = c(\mathcal{X}, \gamma_{min})$.

Proof. Using Theorem 4.3 in Section VII of [13] and Corollary 3.2.8 in [41] for the precise form of the constants, there exist non-negative transition densities p_γ of $(X_s : s \geq 0)$ which satisfy the estimate

$$p_\gamma(t,x,y) \leq c_1 t^{-d/2}, \qquad t > 0, \quad x, y \in \mathbb{R}^d, \tag{6.28}$$

with c_1 depending only on the lower bound $\gamma_{min} \leq \gamma$. Then, arguing as in the proof of Theorem 1.17 in [34], with (6.28) replacing the standard heat kernel estimate for Brownian motion ($\gamma = 1/2$) used in [34], we obtain the desired inequality. □

From what precedes and Jensen's inequality we deduce that if $g = 0, h \geq h_{min} > 0$ then under the maintained assumptions solutions v to (6.6) are bounded away from zero;

$$\inf_{x\in\mathcal{X}} v(x) \geq h_{min} e^{-\|V\|_\infty \sup_x E_\gamma^x \tau_{\mathcal{X}}} \geq h_{min} e^{-c\|V\|_\infty} > 0. \tag{6.29}$$

In a similar vein, for non-negative potentials $V \geq 0$ the solutions v to (6.6) satisfy

$$\|v\|_\infty \leq \|g\|_\infty \sup_{x\in\mathcal{X}} E_\gamma^x \tau_{\mathcal{X}} + \|h\|_\infty \leq c\|g\|_\infty + \|h\|_\infty. \tag{6.30}$$

In particular the solution operator $L_{\gamma,V}^{-1}$ of (6.6) with Dirichlet boundary conditions $h = 0$ is linear and Lipschitz for the $\|\cdot\|_\infty$-norm,

$$\|L_{\gamma,V}^{-1}[g]\|_\infty \leq c\|g\|_\infty, \quad g \in C^\infty(\mathcal{X}), \tag{6.31}$$

with Lipschitz constant c from Lemma 6.1.3, upgrading the L^2- estimate (6.12) to one for the uniform norm.

6.1.5 Elliptic regularity estimates

The Lipschitz estimates (6.12), (6.13), (6.31) for the elliptic solution operator $L^{-1}_{\gamma,V}$ are remarkably 'universal' in the conductivity γ and potential V in that they only depend on γ, V via the lower bounds on these. However, from elliptic PDE theory we expect the solution operator to be 2-smoothing in a natural Sobolev scale. The constants in such stronger estimates depend on higher regularity of the coefficients γ, V and require different proofs.

A starting point is the corresponding result for the standard Laplacian $\Delta = L_{1,0}$ which establishes a linear topological isomorphism

$$u \mapsto (\Delta u, \mathrm{tr}[u]) : H^{k+1}(\mathcal{X}) \to H^{k-1}(\mathcal{X}) \times H^{k+1/2}(\partial\mathcal{X}), \tag{6.32}$$

where $\mathcal{X}$ is a smooth bounded domain in $\mathbb{R}^d$, $\mathrm{tr} = \mathrm{tr}_{\partial\mathcal{X}}$ the boundary trace operator, and integer $k \geq 1$. See Theorem II.5.4 in [87] for a proof. The isomorphism furnishes us with constants k_1, k_2 that depend only on $k, \mathcal{X}, d$ for which

$$k_1\|\Delta u\|_{H^{k-1}} \leq \|u\|_{H^{k+1}} \leq k_2\|\Delta u\|_{H^{k-1}}, \quad \text{if} \;\; u \in H_0^{k+1}(\mathcal{X}), \quad k \in \mathbb{N}. \tag{6.33}$$

Now let us assume $(\gamma, V) \in H^\beta \times H^{\beta-1}$ for $\beta > 1 + d/2$ and write

$$\Delta u = \gamma^{-1}(L_{\gamma,V}u - \nabla\gamma \cdot \nabla u + Vu). \tag{6.34}$$

Then for $g \in L^2(\mathcal{X})$ we have $u = L^{-1}_{\gamma,V}[g] \in H_0^2(\mathcal{X})$ in view of (6.11) with $k = 1$, so the inequality (6.33) gives

$$\begin{aligned} \|L^{-1}_{\gamma,V}[g]\|_{H^2} &\leq k_2\|\Delta L^{-1}_{\gamma,V}[g]\|_{L^2} = k_2\|\gamma^{-1}(g - \nabla\gamma \cdot \nabla L^{-1}_{\gamma,V}[g] + VL^{-1}_{\gamma,V}[g])\|_{L^2} \\ &\leq k_2\gamma_{min}^{-1}(\|g\|_{L^2} + (\|\gamma\|_{C^1} + \|V\|_\infty)\|L^{-1}_{\gamma,V}[g]\|_{H^1}) \\ &\leq C(\gamma_{min}, B)\|g\|_{L^2} \end{aligned} \tag{6.35}$$

where we have used (6.13) in the last line, and where B is an upper bound for $\max(\|\gamma\|_{C^1}, \|V\|_\infty)$.

The 'dual' version of the estimate (6.35) and its converse are as follows.

Proposition 6.1.4. *Suppose $(\gamma, V) \in H^\beta \times H^{\beta-1}$ for some $\beta > 1 + d/2$ and that $\gamma \geq \gamma_{min} > 0$. Then we have for all $g \in L^2(\mathcal{X})$ that*

$$c(B)\|g\|_{(H_0^2)^*} \leq \|L^{-1}_{\gamma,V}[g]\|_{L^2} \leq C(B, \gamma_{min})\|g\|_{(H_0^2)^*}$$

where B is an upper bound for $\max(\|\gamma\|_{C^1}, \|V\|_\infty)$.

Proof. Representing the L^2-norm by duality and using self-adjointness of $L^{-1}_{\gamma,V}$ as well as (6.35) we obtain

$$\begin{aligned}
\|L^{-1}_{\gamma,V}[g]\|_{L^2} &= \sup_{\|\varphi\|_{L^2}\le 1}\left|\int_{\mathcal{X}} L^{-1}_{\gamma,V}[g]\varphi\right| = \sup_{\|\varphi\|_{L^2}\le 1}\left|\int_{\mathcal{X}} gL^{-1}_{\gamma,V}[\varphi]\right| \\
&\le C(\gamma_{min}, B) \sup_{\varphi\in H_0^2,\ \|\varphi\|_{H^2}\le 1}\left|\int_{\mathcal{X}} g\varphi\right| \\
&= C(\gamma_{min}, B)\|g\|_{(H_0^2)^*},
\end{aligned}$$

which completes the proof of the upper bound. For the lower bound, notice first that by the divergence theorem (6.7) and the Cauchy-Schwarz inequality, for any $w \in H_0^2$,

$$\begin{aligned}
\|L_{\gamma,V}w\|_{(H_0^2)^*} &= \sup_{\phi\in H_0^2:\|\phi\|_{H^2}\le 1}\left|\int wL_{\gamma,V}\phi\right| \\
&\le \bar{c}(B)\|w\|_{L^2} \sup_{\phi\in H_0^2:\|\phi\|_{H^2}\le 1}\left[\|\Delta\phi\|_{L^2} + \|\nabla\phi\|_{L^2} + \|\phi\|_{L^2}\right] \\
&\lesssim \bar{c}(B)\|w\|_{L^2}.
\end{aligned}$$

Now in light of (6.35) we can insert $w = L^{-1}_{\gamma,V}[g] \in H_0^2$ in the preceding inequality which yields the desired lower bound. □

It is sometimes convenient to remove the dependence of the constant on B in the upper bound in the previous result, which is possible by considering slightly weaker dual norms: Indeed, replacing (6.35) by (6.13) in the previous proof immediately gives, for $\bar{c} = \bar{c}(\mathcal{X}, \gamma_{min})$,

$$\|L^{-1}_{\gamma,V}[g]\|_{L^2} \le \bar{c}\|g\|_{(H_0^1)^*}. \tag{6.36}$$

A general bound on the Sobolev regularity of solutions of the PDE (6.6) – quantifying (6.11) – is now given in the following proposition. We remark that $g \in C^\infty$ can be relaxed to $g \in H^{\beta-1}$ by simple adjustments to the proof.

Proposition 6.1.5. *Let $\beta > 1 + d/2$. Suppose $v = v_{\gamma,V}$ is the unique solution of (6.6) for $\gamma \in H^\beta(\mathcal{X})$ such that $\gamma \ge \gamma_{min} > 0$, $V \in H^{\beta-1}(\mathcal{X})$ such that $V \ge 0$, and smooth g, h. Then $v \in H^{\beta+1}(\mathcal{X})$ and for every $B > 0$ there exists a constant $C = C(\beta, d, \mathcal{X}, \gamma_{min}, g, h, B) > 0$ such that*

$$\sup_{\|\gamma\|_{H^\beta}+\|V\|_{H^{\beta-1}}\le B} \|v_{\gamma,V}\|_{H^{\beta+1}} \le C < \infty \tag{6.37}$$

Proof. We know that $v \in H^{\beta+1}(\mathcal{X})$ by (6.11) with $k = \beta$ and that

$$c_k(h) := \|h\|_{H^{k+1/2}(\partial\mathcal{X})} < \infty$$

for all k since h is smooth. Then (6.32), (6.34) and the multiplicative inequality (6.2) give

$$\begin{aligned}\|v\|_{H^{\beta+1}} &\lesssim c_\beta(h) + \|\gamma^{-1}(L_{\gamma,V}v - \nabla\gamma\cdot\nabla v + Vv)\|_{H^{\beta-1}}\\ &\lesssim c_\beta(h) + \|\gamma^{-1}\|_{H^{\beta-1}}(\|g\|_{H^{\beta-1}} + (\|\gamma\|_{H^\beta} + \|V\|_{H^{\beta-1}})\|v\|_{H^\beta}).\end{aligned}$$

By the chain rule and (6.2) we can bound

$$\sup_{\|\gamma\|_{H^\beta}\le B, \|V\|_{H^{\beta-1}}\le B} \max\left(\|\gamma^{-1}\|_{H^{\beta-1}}, \|\gamma\|_{H^\beta}, \|V\|_{H^{\beta-1}}\right) \le C(B) < \infty,$$

and then using also the interpolation inequality (6.5) we obtain

$$\|v\|_{H^{\beta+1}} \lesssim c_\beta(h) + \|g\|_{H^{\beta-1}} + \|v\|_{H^\beta} \lesssim c_\beta(h) + \|g\|_{H^{\beta-1}} + \|v\|_{H^{\beta+1}}^{\frac{\beta}{\beta+1}}\|v\|_{L^2}^{\frac{1}{\beta+1}}$$

When $\|v\|_{H^{\beta+1}} \le 1$ we use (6.30) to deduce

$$\|v\|_{H^{\beta+1}} \lesssim c_\beta(h) + \|g\|_{H^{\beta-1}} + \|v\|_\infty^{1/(\beta+1)} \lesssim c_\beta(h) + \|g\|_{H^{\beta-1}} + (\|g\|_\infty + \|h\|_\infty)^{1/(\beta+1)}$$

and when $\|v\|_{H^{\beta+1}} \ge 1$, then dividing both sides by $\|v\|_{H^{\beta+1}}^{\frac{\beta}{\beta+1}}$ and using again (6.30) yields

$$\|v\|_{H^{\beta+1}}^{1/(\beta+1)} \lesssim c_\beta(h) + \|g\|_{H^{\beta-1}} + \|v\|_\infty^{1/(\beta+1)} \le c_\beta(h) + \|g\|_{H^{\beta-1}} + (\|g\|_\infty + \|h\|_\infty)^{1/(\beta+1)}.$$

Raising the last inequality to the power $\beta + 1$ and taking maxima from the last two inequalities implies (6.37). □

We finally give an explicit Lipschitz estimate for the solution operator $L_{\gamma,V}^{-1}$ between general Sobolev spaces.

Proposition 6.1.6. *Suppose $0 < \gamma_{min} \le \gamma$ and that $\|\gamma\|_{C^\beta} + \|V\|_{C^{\beta-1}} \le B$ for some $\beta \ge 1$. Then for some $C = C(B, \beta, d, \mathcal{X}, \gamma_{min})$ we have for all $g \in H^{\beta-1}$ that*

$$\|L_{\gamma,V}^{-1}[g]\|_{H^{\beta+1}} \le C\|g\|_{H^{\beta-1}}.$$

Proof. By (6.11) we know that $L_{\gamma,V}^{-1}[g] \in H_0^{\beta+1}$. From (6.33) and (6.34) we see for any $u \in H_0^{\beta+1}$ that

$$\|u\|_{H^{\beta+1}} \le k_2\|\Delta u\|_{H^{\beta-1}} \le \|\gamma_{min}^{-1}\|_{C^{\beta-1}}\|L_{\gamma,V}u\|_{H^{\beta-1}} + \|\gamma\|_{C^\beta}\|u\|_{H^\beta} + \|V\|_{C^{\beta-1}}\|u\|_{H^{\beta-1}}.$$

By the hypotheses, $\|\gamma_{min}^{-1}\|_{C^{\beta-1}} + \|\gamma\|_{C^\beta} + \|V\|_{C^{\beta-1}}$ is bounded by a fixed constant, B' say. We will show that the previous inequality implies that

$$\|u\|_{H^{\beta+1}} \leq C\|L_{\gamma,V}u\|_{H^{\beta-1}} \quad \forall u \in H_0^{\beta+1}, \tag{6.38}$$

which then also implies the proposition by taking $u = L_{\gamma,V}^{-1}[g]$. To prove (6.38) suppose by way of contradiction that the inequality is not true. Then there exists a sequence $u_m \in H_0^{\beta+1}$ such that $\|u_m\|_{H^{\beta+1}} = 1$ but $\|L_{\gamma,V}u_m\|_{H^{\beta-1}} \to 0$ as $m \to \infty$. By compactness of the imbedding $H_0^{\beta+1} \subset H_0^\beta \subset H_0^1 = H_c^1$ (the latter being a Banach space), this sequence converges (if necessary by passing to a subsequence) to some $u \in H_0^1$ satisfying $L_{\gamma,V}u = 0$. The last but one displayed inequality implies that the sequence u_m is Cauchy in $H^{\beta+1}$ and hence its limit u_0 must also satisfy $\|u_0\|_{H^{\beta+1}} = 1$, in particular $u_0 \neq 0$. But from the results from Section 6.1.2 we know that the unique solution in H_0^1 to $L_{\gamma,V}u = 0$ equals $u_0 = 0$, a contradiction. □

6.2 Further auxiliary results

6.2.1 Results from Gaussian process theory

A Borel probability measure μ on a separable Banach space B is called *centred Gaussian* if the image measures $\mu \circ L^{-1}$ are mean zero normal distributions on $\mathbb{R}$ for all L in the topological dual space B^* of B. A (Borel-) random variable X defined on some probability space $(\Omega, \mathcal{A}, \Pr)$ taking values in a separable Banach space $(B, \|\cdot\|_B)$ is called centred Gaussian if its law $\mu_X = \Pr \circ X^{-1}$ defines a centred Gaussian probability measure on B.

A *centred Gaussian process* with index set T is a collection $\{X(t) : t \in T\}$ of random variables such that any finite selection $\{X(t_1), \dots, X(t_d)\}, t_i \in T, d \in \mathbb{N}$, is multivariate centred normal in $\mathbb{R}^d$.

Any Gaussian random variable in a Banach space can be viewed as a Gaussian process indexed by B^*, and conversely often Gaussian processes can be realised as B-valued random variables for appropriate B, e.g., Brownian motion $(X(t) : t \in [0,1])$ is a Gaussian process which (by sample-path continuity) also defines a Gaussian random variable in the space of continuous functions $C([0,1])$. So the theories of Gaussian measures and processes inform each other and tools from both fields are useful when deployed appropriately. See Section 2 in [61] for an account of the basic theory and many references.

There are some classical results about properties of Gaussian measures and processes, such as *Fernique's theorem*, the *Karhunen-Loève expansion*, the *Sudakov-Cirelson* and *log-Sobolev inequalities*, *Borell's isoperimetric* and the *Cameron-Martin theorem.* The last two theorems involve the notion of a *reproducing kernel*

Hilbert space (RKHS) $\mathcal{H}$ of a Gaussian measure μ whose inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ encodes the geometry of the covariance structure of Gaussian variables $X \sim \mu$. We recall that $\mathcal{H}$ is a subspace of B that is compactly embedded in the sense that the identity map

$$id : (\mathcal{H}, \| \cdot \|_{\mathcal{H}}) \to (B, \| \cdot \|_B)$$

is compact (in particular continuous). One can also define a 'process-RKHS' for Gaussian processes (instead of for measures), and when the law of the process defines a Gaussian probability measure on B, these RKHS will coincide under mild regularity conditions (e.g., Lemma 11.14 in [59]). We will admit these facts and theorems without stating them here, even though we will give precise references to the literature when they are used in our proofs.

There are a couple of results on Gaussian measures that are not covered in Chapter 2 in [61] that we review here now.

Small deviation estimates for Gaussian measures

The following theorem characterises the 'small ball' (or 'small deviation') asymptotics of a Gaussian measure μ on a separable Banach space B quantified by the function

$$\phi(\varepsilon) = -\log \mu\big(x \in B : \|x\|_B \le \varepsilon\big), \quad \varepsilon > 0, \tag{6.39}$$

in terms of the metric entropy function

$$\mathscr{H}(\varepsilon) = \log N\big(\{x \in B : \|x\|_{\mathcal{H}} \le 1\}, \| \cdot \|_B, \varepsilon\big), \quad \varepsilon > 0, \tag{6.40}$$

which equals the logarithm of the minimal number of B-balls of radius $\varepsilon > 0$ required to cover the (compact) unit ball of $\mathcal{H} \subset B$.

Theorem 6.2.1 (Kuelbs-Li-Linde)**.** *Let $a > 0$. Then as $\varepsilon \to 0$,*

$$\phi(\varepsilon) \approx \varepsilon^{-a} \tag{6.41}$$

if and only if

$$\mathscr{H}(\varepsilon) \approx \varepsilon^{-\frac{2a}{2+a}}, \tag{6.42}$$

where $\approx$ denotes two-sided inequalities up to multiplicative constants (that may depend on B).

For a proof of this theorem, see Theorem 1.2 in [85] with $\alpha = a$ and $\beta = 0$, with a key earlier reference being [78]. A refinement of this result at the 'log-scale' ($\beta > 0$ in Theorem 1.2 in [85]) is possible but not required in the present context. We also refer to [86] for more on the topic of small ball deviation estimates for Gaussian processes.

The Gaussian correlation inequality

A result in the theory of Gaussian processes that has been proved only recently is the Gaussian correlation inequality (formerly conjecture!) – the original proof of Royen is spelt out in [80].

Theorem 6.2.2 (Gaussian correlation inequality)**.** *Let μ be a centred Gaussian measure on a separable Banach space B and let C_1, C_2 be convex closed symmetric subsets of B. Then*

$$\mu(C_1 \cap C_2) \geq \mu(C_1)\mu(C_2). \tag{6.43}$$

A proof for finite-dimensional Gaussian distributions ($dim(B) < \infty$) is given in [80]. As the conclusion of the result does not depend on $dim(B)$, the Karhunen-Loève theorem combined with a standard limiting argument extends the result to infinite-dimensions, see Appendix A.4 of [64] for details.

Whittle-Matérn-type Gaussian processes

We can construct a centred Gaussian process $X = \{X(z),\ z \in \mathcal{Z}\}$ indexed by an arbitrary subset $\mathcal{Z} \subset \mathbb{R}^d$ by prescribing a *positive definite function* (covariance kernel) $K : \mathbb{R}^d \to \mathbb{R}$ such that $EX(s)X(t) = K(s-t), s, t \in \mathcal{Z}$, see Proposition 2.1.10 in [61]. One such choice (with $|\cdot| = \sqrt{\langle \cdot, \cdot \rangle_{\mathbb{R}^d}}$ the Euclidean norm) is

$$K_\alpha(x) = \int_{\mathbb{R}^d} e^{i\langle x, \xi \rangle_{\mathbb{R}^d}} d\bar{\mu}(\xi), \quad d\bar{\mu}(\xi) = (1 + |\xi|^2)^{-\alpha} d\xi, \quad x \in \mathcal{Z},$$

whenever $\alpha > d/2$. That K_α is positive definite is the easy part of 'Bochner's theorem': For any finite collection $\beta_j \in \mathbb{C}, s_j \in \mathcal{Z}$, we have

$$\begin{aligned} \sum_{k,j} \beta_k \overline{\beta_j} K_\alpha(s_k - s_j) &= \int_{\mathbb{R}^d} \sum_k \beta_k e^{i\langle s_k, \xi \rangle_{\mathbb{R}^d}} \sum_j \overline{\beta_j e^{i\langle s_j, \xi \rangle_{\mathbb{R}^d}}} d\bar{\mu}(\xi) \\ &= \int_{\mathbb{R}^d} \Big| \sum_j \beta_j e^{i\langle s_j, \xi \rangle_{\mathbb{R}^d}} \Big|^2 d\bar{\mu}(\xi) \geq 0. \end{aligned}$$

Such (stationary) Gaussian process priors are popular in numerical analysis, spatial statistics and machine learning, and are sometimes called 'Whittle-Matérn' processes. The following theorem permits their use in the setting of these notes.

Theorem 6.2.3 (Existence of Whittle-Matérn prior measures)**.** *Let $\mathcal{Z}$ be a bounded domain in $\mathbb{R}^d$ with smooth boundary. Then the law of the Gaussian process $X = \{X(z),\ z \in \mathcal{Z}\}$ with covariance function $K_\alpha, \alpha > d/2$, defines a centred Gaussian probability measure μ_X on the Banach space $C_u(\mathcal{Z})$ of bounded uniformly continuous function on $\mathcal{Z}$ with reproducing kernel Hilbert space $\mathcal{H} = H^\alpha(\mathcal{Z})$. We further have $\mu_X(H^\beta(\mathcal{Z})) = 1$ for every $0 \leq \beta < \alpha - d/2$.*

Proof. The process-RKHS $\mathcal{H}_{\alpha,\mathcal{Z}}$ of $\{X(z) : z \in \mathcal{Z}\}$ can be characterised as the restriction to $\mathcal{Z}$ of functions in

$$\mathcal{H}_{\bar{\mu}} = \Big\{ g : \mathbb{R}^d \to \mathbb{R} \text{ Borel meas. } : \|g\|^2_{\mathcal{H}_{\bar{\mu}}} \equiv \int_{\mathbb{R}^d} |\hat{g}|^2 d\bar{\mu} < \infty \Big\},$$

with $\hat{g} = \int_{\mathbb{R}^d} e^{i\langle \cdot, x\rangle_{\mathbb{R}^d}} g(x)dx$ the Fourier transform of g, equipped with the quotient norm

$$\|h\|_{\mathcal{H}_{\alpha,\mathcal{Z}}} = \inf_{g \in \mathcal{H}_{\bar{\mu}}, g=h \text{ on } \mathcal{Z}} \|g\|_{\mathcal{H}_{\bar{\mu}}},$$

see Lemma 11.35 in [59] for a proof of this fact. Then by, e.g., Theorem 1.9.2 in [87] and for $\mathcal{Z}$ a bounded smooth domain in $\mathbb{R}^d$, the norm of $\mathcal{H}_{\alpha,\mathcal{Z}}$ is Lipschitz equivalent to the classically defined Sobolev norm (cf. (6.1)) of $H^\alpha(\mathcal{Z})$, and hence the process RKHS of $\{X(z),\ z \in \mathcal{Z}\}$ coincides with $H^\alpha(\mathcal{Z})$ for such $\mathcal{Z}$.

The Gaussian process $(X(z) : z \in \mathcal{Z})$ has intrinsic covariance metric

$$d_X(s,t) = \sqrt{E(X(s) - X(t))^2}, \quad s,t \in \mathcal{Z},$$

satisfying for some $0 < \gamma < \alpha - d/2$ the estimate

$$d_X^2(s,t) = EX^2(s) + EX^2(t) - 2EX(s)X(t) = 2(K_\alpha(0) - K_\alpha(s-t)) \le c|s-t|^\gamma$$

because $e^{-i\langle \cdot, \xi\rangle_{\mathbb{R}^d}}$ is γ-Hölder with Hölder constant $|\xi|^\gamma$ and since $(1+|\xi|^2)^{-\alpha}|\xi|^\gamma$ is integrable on $\mathbb{R}^d$. This implies (using, e.g., Proposition 4.3.34 in [61]) that the bounded subset $\mathcal{Z}$ of $\mathbb{R}^d$ can be covered by at most $N(\epsilon) = (A/\epsilon)^{2d/\gamma}$ balls of d_X-radius $\epsilon > 0$. Since $\sqrt{\log N(\epsilon)}$ is $d\epsilon$-integrable at zero, Theorem 2.3.7 in [61] implies that there exists a version of X that is uniformly continuous on $\mathcal{Z}$ for the d_X and then also the $|\cdot|$-metric. By Proposition 2.1.5 in [61] the law of X then defines a Gaussian Borel probability measure μ_X on the separable Banach space $C_u(\mathcal{Z})$ of bounded uniformly continuous functions on $\mathcal{Z}$, and the process RKHS $\mathcal{H}_{\alpha,\mathcal{Z}}$ of $(X(z) : z \in \mathcal{Z})$ co-incides with the RKHS $\mathcal{H}$ of μ_X by virtue of Lemma 11.14 in [59].

To see that $\mu_X(\|X\|_{H^\beta} < \infty) = 1$ whenever $0 < \beta < \alpha - d/2$, use the Karhunen-Loéve theorem 2.6.10 in [61] to represent $X = \sum_i g_i h_i$ almost surely in $C(\mathcal{Z})$, where the (h_i) form an orthonormal basis of the RKHS $\mathcal{H}^\alpha(\mathcal{Z})$, and the g_i are iid standard normal $N(0,1)$ variables. The identity imbedding map $H^\alpha(\mathcal{Z}) \to H^\beta(\mathcal{Z})$ is Hilbert-Schmidt for any $\beta < \alpha - d/2$, in particular

$$\sum_{i=1}^\infty \|h_i\|^2_{H^\beta} < \infty, \tag{6.44}$$

see Theorem 6.61 in [2]. By the theory of radonifying maps this implies $\Pr(X \in H^\beta) = 1$, see Theorem 3.20 and Proposition 13.5 in [131]. $\square$

We note that the previous proof shows that the theorem remains valid for arbitrary open bounded subsets $\mathcal{Z}$ (possibly with irregular boundary) of $\mathbb{R}^d$ as long as $\mathcal{H}$ is taken to equal the space $\mathcal{H}_{\alpha,\mathcal{Z}}$ introduced in the proof.

One can obtain stronger (than H^β) almost sure Hölder regularity properties of the random function $z \mapsto X(z)$ by more refined arguments (e.g., Proposition I.4 in [59] or p.22f. in [3]). In these notes we confine ourselves to work with Sobolev-regularity – this combines nicely with regularity estimates for elliptic PDEs used elsewhere and also permits a 'soft' proof of path regularity using only the Karhunen-Loeve theorem and Hilbert-Schmidt embeddings.

The preceding arguments show that for any $\alpha > d/2$ we can find a Gaussian prior $\Pi = \mu_X$ that has $H^\alpha(\mathcal{Z})$ as RKHS. In some of the results in these notes we wish to enforce a certain boundary behaviour of the prior Π. We can achieve this by taking a cutoff-function $\xi \in C^\infty(\mathcal{Z})$ of compact support in $\mathcal{Z}$ and consider a new Gaussian process $(X_\xi(z) = X(z)\xi(z) : z \in \mathcal{Z})$. The RKHS of this new process then equals $\mathcal{H}_\xi = \{g = \xi h : h \in H^\alpha(\mathcal{Z})\} \subset H^\alpha_c(\mathcal{Z})$ with norm $\|\xi^{-1} \cdot \|_{H^\alpha}$. We can also take a different approach by starting with a Gaussian series expansion

$$X(z) = \sum_{j=1}^{\infty} g_j \lambda_j^{-\alpha/2} e_j(z), \ z \in \mathcal{Z}, \tag{6.45}$$

of the eigenfunctions of the Dirichlet Laplacian $L_{1,0}$ from (6.9). Their laws define Gaussian probability measures on $C(\mathcal{Z})$ and as in Example 2.6.15 in [61] one shows that their RKHS are the closed subspaces $\tilde{H}^\alpha(\mathcal{Z}) \subset H^\alpha_0(\mathcal{Z})$ of $H^\alpha(\mathcal{Z})$ from (6.15).

6.2.2 A concentration inequality for empirical processes

Inequalities for suprema of empirical process were used at several places in these notes. These results have a long history (see Chapter 3 in [61]) and are particularly well developed for uniformly bounded variables. In the random design setting here, the processes are often not uniformly bounded due to the presence of the Gaussian noise variables ε_i. This can be dealt with as in the proof of Lemma 4.1.7 under integrability conditions only on the 'square-root' metric entropy in (4.6). However, if one is willing to work with stronger hypothesis of integrability of the metric entropy (without the square-root), a more basic concentration argument is sometimes equally useful. The following lemma is based on a chaining argument for stochastic processes with a mixed tail (cf. Theorem 2.2.28 in Talagrand [121] and also Theorem 3.5 in [45]). For us it will be sufficient to control the 'generic chaining' functionals employed in these references by suitable metric entropy integrals. For any (semi-)metric d on a metric space T, we denote by $N = N(T, d, \rho)$ the minimal cardinality of a covering of T by balls with centres $(t_i : i = 1, \dots, N) \subset T$ such that for all $t \in T$ there exists i such that $d(t, t_i) < \rho$. Below we require the index set Θ to be countable (to avoid measurability issues). Whenever we apply Lemma 6.2.4 in these notes with an uncountable set Θ, one can show that the supremum can be realised as one over a countable subset of it.

Lemma 6.2.4. *Let Θ be a countable set. Suppose a class of real-valued measurable functions*

$$\mathcal{H} = \{h_\theta : \mathcal{X} \to \mathbb{R}, \theta \in \Theta\}$$

defined on a probability space $(\mathcal{X}, \mathcal{A}, P^X)$ is uniformly bounded by $U \geq \sup_\theta \|h_\theta\|_\infty$ and has variance envelope $\sigma^2 \geq \sup_\theta E^X h_\theta^2(X)$ where $X \sim P^X$. Define metric entropy integrals

$$J_2(\mathcal{H}) = \int_0^{4\sigma} \sqrt{\log N(\mathcal{H}, d_2, \rho)} d\rho, \quad d_2(\theta, \theta') := \sqrt{E^X[h_\theta(X) - h_{\theta'}(X)]^2},$$

$$J_\infty(\mathcal{H}) = \int_0^{4U} \log N(\mathcal{H}, d_\infty, \rho) d\rho, \quad d_\infty(\theta, \theta') := \|h_\theta - h_{\theta'}\|_\infty.$$

For $X_1, \dots, X_N$ drawn i.i.d. from P^X and $\varepsilon_i \sim^{iid} N(0,1)$ independent of all the X_i's, consider empirical processes arising either as

$$Z_N(\theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^N h_\theta(X_i)\varepsilon_i, \quad \theta \in \Theta,$$

or as

$$Z_N(\theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^N (h_\theta(X_i) - Eh_\theta(X)), \quad \theta \in \Theta.$$

We then have for some universal constant $L > 0$ and all $x \geq 1$,

$$\Pr\left(\sup_{\theta \in \Theta} |Z_N(\theta)| \geq L\Big[J_2(\mathcal{H}) + \sigma\sqrt{x} + (J_\infty(\mathcal{H}) + Ux)/\sqrt{N}\Big]\right) \leq 2e^{-x}.$$

Proof. We only prove the case where $Z_N(\theta) = \sum_i h_\theta(X_i)\varepsilon_i/\sqrt{N}$, the simpler case without Gaussian multipliers is proved in the same way. We will apply Theorem 3.5 in [45], whose condition (3.8) we need to verify. First notice that for $|\lambda| < 1/\|h_\theta - h_{\theta'}\|_\infty$, and E^ε denoting the expectation with respect to ε,

$$\begin{aligned} E\exp\left\{\lambda\varepsilon(h_\theta - h_{\theta'})(X)\right\} &\leq 1 + \sum_{k=2}^\infty \frac{|\lambda|^k E^\varepsilon|\varepsilon|^k E^X|h_\theta - h_{\theta'}|^k(X)}{k!} \\ &\leq 1 + \lambda^2 E^X[h_\theta(X) - h_{\theta'}(X)]^2 \sum_{k=2}^\infty \frac{E^\varepsilon|\varepsilon|^k}{k!}\big(|\lambda|\|h_\theta - h_{\theta'}\|_\infty\big)^{k-2} \\ &\leq \exp\left\{\frac{\lambda^2 d_2^2(\theta,\theta')}{1 - |\lambda| d_\infty(\theta,\theta')}\right\} \qquad (6.46) \end{aligned}$$

where we have used the basic fact $E^\varepsilon|\varepsilon|^k/k! \leq 1$. By the i.i.d. hypothesis we then also have

$$E\exp\Big\{\lambda(Z_N(\theta)-Z_N(\theta'))\Big\} \leq \exp\left\{\frac{\lambda^2 d_2^2(\theta,\theta')}{1-|\lambda| d_\infty(\theta,\theta')/\sqrt{N}}\right\}.$$

An application of the exponential Chebyshev inequality (and optimisation in λ, as in the proof of Proposition 3.1.8 in [61]) then implies that condition (3.8) in [45] holds for the stochastic process $Z_N(\theta)$ with metrics $\bar{d}_2 = 2d_2$ and $\bar{d}_1 = d_\infty/\sqrt{N}$. In particular, the $\bar{d}_2$-diameter $\Delta_2(\mathcal{H})$ of $\mathcal{H}$ is at most 4σ and the $\bar{d}_1$-diameter $\Delta_1(\mathcal{H})$ of $\mathcal{H}$ is bounded by $4U/\sqrt{N}$. [These bounds are chosen so that they remain valid for the process without Gaussian multipliers as well.] Theorem 3.5 in [45] now gives, for some universal constant M, and any $\theta_\dagger \in \Theta$ that

$$\Pr\left(\sup_{\theta\in\Theta}|Z_N(\theta)-Z_N(\theta_\dagger)| \geq M\big(\gamma_2(\mathcal{H})+\gamma_1(\mathcal{H})+\sigma\sqrt{x}+(U/\sqrt{N})x\big)\right) \leq e^{-x}$$

where the 'generic chaining' functionals γ_1, γ_2 are upper bounded by the respective metric entropy integrals of the metric spaces $(\mathcal{H}, \bar{d}_i), i = 1, 2$, up to universal constants (see (2.3) in [45]). For γ_1 also notice that a simple substitution $\rho' = \rho\sqrt{N}$ implies that

$$\int_0^{4U/\sqrt{N}} \log N(\mathcal{H}, \bar{d}_1, \rho)d\rho = \frac{1}{\sqrt{N}}\int_0^{4U} \log N(\mathcal{H}, d_\infty, \rho')d\rho',$$

and we hence deduce that

$$\Pr\left(\sup_{\theta\in\Theta}|Z_N(\theta)-Z_N(\theta_\dagger)| \geq \bar{L}\Big[J_2(\mathcal{H})+\sigma\sqrt{x}+(J_\infty(\mathcal{H})+Ux)/\sqrt{N}\Big]\right) \leq e^{-x} \tag{6.47}$$

for some universal constant $\bar{L}$.

Now what precedes also implies the classical Bernstein-inequality

$$\Pr\left(|Z_N(\theta)| \geq \sigma\sqrt{2x}+\frac{Ux}{3\sqrt{N}}\right) \leq 2e^{-x},\ x>0, \tag{6.48}$$

for any fixed $\theta \in \Theta, U \geq \|h_\theta\|_\infty$ and $\sigma^2 \geq E^X h_\theta^2(X)$, proved as (3.24) in [61], using (6.46). Applying this with $\theta_\dagger$ and using (6.47), the final result follows now from

$$\begin{aligned}\Pr\big(\sup_{\theta\in\Theta}|Z_N(\theta)| > 2\tau(x)\big) &\leq \Pr\big(\sup_{\theta\in\Theta}|Z_N(\theta)-Z_N(\theta_\dagger)| > \tau(x)\big)\\ &\quad + \Pr\big(|Z_N(\theta_\dagger)| > \tau(x))\big) \leq 2e^{-x},\end{aligned}$$

for any $x \geq 1$, where $\tau(x) = \bar{L}\big[J_2(\mathcal{H})+\sigma\sqrt{x}+(J_\infty(\mathcal{H})+Ux)/\sqrt{N}\big]$ and $L \geq 2\bar{L} > 0$ is large enough. □

6.2.3 Mixing time bounds for Langevin diffusions

In this section we collect some results about convergence guarantees for an Unadjusted Langevin Algorithm (ULA) from (1.21) for sampling from *strongly log-concave target probability measures*. Early esults of this kind are due to [114] but the recent focus has been on non-asymptotic results in high-dimensional settings, see [39, 49, 50]. Some of the ideas underpinning the fast mixing of 'hypercontractive' Langevin diffusions (1.20) in high-dimensions go back a long way further, we mention [6], [73], the recent monograph [7], and further references below

Suppose that μ is a Borel probability measure on $\mathbb{R}^D$ which has a Lebesgue density proportional to e^{-U} for some potential $U : \mathbb{R}^D \to \mathbb{R}$, specifically

$$\mu(B) = \frac{\int_B e^{-U(\theta)} d\theta}{\int_{\mathbb{R}^D} e^{-U(\theta)} d\theta}, \quad B \subseteq \mathbb{R}^D \text{ measurable.} \tag{6.49}$$

Following [50] (cf. H1 and H2 there) we will assume that the potential U has a Λ-Lipschitz gradient and is m-strongly convex, in particular μ is then log-concave.

Assumption 6.2.5. *1. The function $U : \mathbb{R}^D \to \mathbb{R}$ is continuously differentiable and there exists a constant $\Lambda \geq 0$ such that for all $\theta, \bar{\theta} \in \mathbb{R}^D$,*

$$\|\nabla U(\theta) - \nabla U(\bar{\theta})\|_{\mathbb{R}^D} \leq \Lambda \|\theta - \bar{\theta}\|_{\mathbb{R}^D}.$$

2. There exists a constant $0 < m \leq \Lambda$ such that for all $\theta, \bar{\theta} \in \mathbb{R}^D$, we have

$$U(\bar{\theta}) \geq U(\theta) + \langle \nabla U(\theta), \bar{\theta} - \theta \rangle_{\mathbb{R}^D} + \frac{m}{2} \|\theta - \bar{\theta}\|^2_{\mathbb{R}^D}.$$

Under Assumption 6.2.5, the potential U has a unique minimiser over $\mathbb{R}^D$, which we shall denote by θ_U.

Now if $(L_t : t \geq 0)$ is the D-dimensional Langevin Markov diffusion process from (1.20) with potential U satisfying Assumption 6.2.5, and if $P_t^{\theta_{\text{init}}}$ is its distribution at time t when started at $\theta_{\text{init}} \in \mathbb{R}^D$, then one can show a mixing time inequality, with W_2 from (5.15) with $E_D = \mathbb{R}^D$,

$$W_2(P_t^{\theta_{\text{init}}}, \mu) \leq e^{-mt} \big[\|\theta_{\text{init}} - \theta_U\|_{\mathbb{R}^D} + \sqrt{D/m} \big], \tag{6.50}$$

which exhibits only a weak dependence on dimension D. A direct proof of (6.50) based on coupling ideas [33] can be found in Proposition 1 in [50]. A functional analytic interpretation is that log-concave measures μ satisfy a $(1/m)$ - log-Sobolev inequality (Prop. 5.7.1 in [7]) and the resulting dimension-free mixing bound for the entropic distance (p.244 in [7]) carries over to the W_2-distance by the quadratic transportation cost inequality of Otto-Villani (Theorem 9.6.1 in [7]).

Now recall the Euler-Maruyama discretisation (1.21) of the Markov diffusion process (1.20) giving rise to the discrete-time Markov chain $(\vartheta_k : k \geq 0)$. We denote by $\mathbf{P}_{\theta_{\text{init}}}, \mathbf{E}_{\theta_{\text{init}}}$ the law and expectation operator, respectively, of the Markov chain $(\vartheta_k : k \geq 1)$ when started at a deterministic point $\vartheta_0 = \theta_{\text{init}}$. We also write $\mathcal{L}(\vartheta_k)$ for the (marginal) distribution of the k-th iterate ϑ_k. The following result now gives a 'discrete' analogue of (6.50).

Proposition 6.2.6. *Suppose that U satisfies Assumption 6.2.5. Then we have for μ as in (6.49) that*

$$W_2^2(\mathcal{L}(\vartheta_k), \mu) \leq 2\big(1 - \frac{m\gamma}{2}\big)^k \Big[\|\theta_{\text{init}} - \theta_U\|^2_{\mathbb{R}^D} + \frac{D}{m}\Big] + \frac{b(\gamma)}{2}, \qquad k \geq 0, \tag{6.51}$$

where

$$b(\gamma) = 36\frac{\gamma D \Lambda^2}{m^2} + 12\frac{\gamma^2 D \Lambda^4}{m^3}, \tag{6.52}$$

Proof. The bound (6.51) follows from an application of Theorem 5 in [50] with fixed step size $\gamma > 0$, where in our case, noting again that $\kappa \in [m, 2m]$, the expression $u_n^{(1)}(\gamma)$ there is upper bounded by $2\big(1 - m\gamma/2\big)^k$ and the expression $u_n^{(2)}(\gamma)$ there is upper bounded by (using that $\gamma \leq \min\{2/\Lambda, 1/m\} \leq \min\{2/\Lambda, 2/\kappa\}$)

$$\begin{aligned}
&\Lambda^2 D\gamma^2\big(\kappa^{-1} + \gamma\big)\Big(2 + \frac{\Lambda^2\gamma}{m} + \frac{\Lambda^2\gamma^2}{6}\Big)\sum_{i=1}^{k}(1 - \kappa\gamma/2)^{k-i} \\
&\leq \Lambda^2 D\gamma^2\big(\kappa^{-1} + \gamma\big)\Big(2 + \frac{\Lambda^2\gamma}{m} + \frac{\Lambda^2\gamma^2}{6}\Big)\frac{2}{\kappa\gamma} \\
&\leq \Lambda^2 D\gamma\Big(\kappa^{-2} + \frac{\gamma}{\kappa}\Big)\Big(6 + \frac{2\Lambda^2\gamma}{m}\Big) \\
&\leq \Lambda^2 D\gamma m^{-2}\Big(18 + \frac{6\Lambda^2\gamma}{m}\Big),
\end{aligned}$$

which equals (6.52). □

For any measurable function $H : \mathbb{R}^D \to \mathbb{R}$ and any $J_{in}, J \geq 0$, let us define the average of H along an ULA trajectory after 'burn-in' period J_{in} by

$$\hat{\mu}^J_{J_{in}}(H) = \frac{1}{J}\sum_{k=J_{in}+1}^{J_{in}+J} H(\vartheta_k).$$

Proposition 6.2.7. *Suppose that U satisfies Assumption 6.2.5 and suppose $\gamma \leq 2/(m+\Lambda)$. Then for all $J, J_{in} \geq 1, x > 0$ and any Lipschitz function $H : \mathbb{R}^D \to \mathbb{R}$, we have the concentration inequality*

$$\mathbf{P}_{\theta_{\text{init}}}\Big(\hat{\mu}^J_{J_{in}}(H) - \mathbf{E}_{\theta_{\text{init}}}[\hat{\mu}^J_{J_{in}}(H)] \geq x\Big) \leq \exp\Big(-\frac{J\gamma x^2 m^2}{16\|H\|^2_{Lip}(1 + 2/(mJ\gamma))}\Big).$$

Moreover

$$\Big(\mathbf{E}_{\theta_{\text{init}}}[\hat{\mu}^J_{J_{in}}(H)] - E_\mu H\Big)^2 \le \|H\|^2_{Lip}\ \frac{1}{J}\sum_{k=J_{in}+1}^{J_{in}+J} W_2^2(\mathcal{L}(\vartheta_k),\mu). \tag{6.53}$$

Proof. The first statement follows from Theorem 17 of [50], noting that $\kappa = 2m\Lambda/(m+\Lambda) \in [m, 2m]$ and that the constant $v_{N,n}(\gamma)$ from (28) of [50] can be upper bounded by

$$1 + \frac{m^{-1} + 2/(m+\Lambda)}{\gamma J} \le 1 + \frac{2}{m\gamma J}.$$

The display (6.53) is derived in (27) of [50]. □

6.2.4 A characterisation of vanishing efficient information

The following theorem is due to [126].

Theorem 6.2.8. *In the setting of Theorem 3.1.5, assume that ψ satisfies (3.20). Then $i_{\theta,H,\psi} = 0$.*

Proof. We prove the result for the case when the information operator $\mathbb{I}_\theta^*\mathbb{I}_\theta$ is *compact* on the Hilbert space $\bar{H} = \overline{(H, \langle\cdot,\cdot\rangle_{L^2_\lambda})}$, relevant in the setting of inverse problems in these notes. [The general result is proved in [126], Theorem 4.1.]

Let us write $I \equiv \mathbb{I}_\theta, L^2 = L^2_\lambda(\mathcal{X})$ in this proof, and let $\ker(I^*I) = \{h \in \bar{H} : I^*Ih = 0\}$. If I^*I is a compact operator on $\bar{H}$ then by the spectral theorem for self-adjoint operators, there exists an orthonormal system of $\bar{H}$ of eigenvectors $\{e_k : k \in \mathbb{N}\}$ spanning $\bar{H} \ominus \ker(I^*I)$ corresponding to eigenvalues $\lambda_k > 0$ so that

$$I^*Ie_k = \lambda_k e_k, \quad \text{and } I^*Ih = \sum_k \lambda_k \langle h, e_k\rangle_{L^2_\zeta} e_k, \quad h \in \bar{H}.$$

We can then define the usual square root operator $(I^*I)^{1/2}$ by

$$(I^*I)^{1/2}h = \sum_k \lambda_k^{1/2}\langle h, e_k\rangle_{L^2_\zeta} e_k, \quad h \in \bar{H}. \tag{6.54}$$

If we denote by P_0 the L^2_ζ-projection onto $\ker(I^*I)$, then the range of $(I^*I)^{1/2}$ equals

$$R((I^*I)^{1/2}) = \Big\{g \in \bar{H} : P_0(g) = 0, \sum_k \lambda_k^{-1}\langle e_k, g\rangle^2_{L^2_\zeta} < \infty\Big\}. \tag{6.55}$$

Indeed using standard Hilbert space arguments, a) since $P_0(e_k) = 0$ for all k, for any $h \in \bar{H}$ the element $g = (I^*I)^{1/2}h$ belongs to the right hand side in the last display, and conversely b) if g satisfies $P_0(g) = 0$ and $\sum_k \lambda_k^{-1} \langle e_k, g\rangle^2_{L^2_\zeta} < \infty$ then $h = \sum_k \lambda_k^{-1/2} \langle e_k, g\rangle e_k$ belongs to $\bar{H}$ and $(I^*I)^{1/2}h = g$.

Next, Lemma A.3 in [126] implies that $R(I^*) = R((I^*I)^{1/2})$. Now suppose $\psi \in \bar{H}$ is such that $\psi \notin R(I^*)$ and hence $\psi \notin R((I^*I)^{1/2})$. Then from (6.55), either $P_0(\psi) \neq 0$ or $\sum_k \lambda_k^{-1}\langle e_k, \psi\rangle^2_{L^2_\zeta} = \infty$ (or both). In the first case, let $\bar{h} = P_0(\psi)$ so

$$\|I\bar{h}\|_{L^2} = \|I(P_0(\psi))\|_{L^2} = \langle I^*I(P_0(\psi)), P_0(\psi)\rangle_{L^2_\zeta} = 0$$

but $\langle \psi, \bar{h}\rangle_{L^2_\zeta} = \|P_0\psi\|^2_{L^2_\zeta} = \delta$ for some $\delta > 0$. Since H is dense in $\bar{H}$, for any $\epsilon, 0 < \epsilon < \min(\delta/(2\|\psi\|_{L^2_\zeta}), \delta^2/4)$, we can find $h \in H$ such that $\|h - \bar{h}\|_{L^2_\zeta} < \epsilon$ and by continuity also $\|I(h - \bar{h})\|_{L^2} < \epsilon$. Then

$$\sqrt{i_{\theta,h,\psi}} = \frac{\|Ih\|_{L^2}}{|\langle \psi, h\rangle_{L^2_\zeta}|} \leq 2\frac{\epsilon}{\delta} \leq \sqrt{\epsilon}.$$

We conclude that $i_{\theta,H,\psi} < \epsilon$ in (3.18), so that the result follows since ϵ was arbitrary. In the second case we have $\sum_k \lambda_k^{-1}\langle e_k, \psi\rangle^2_{L^2_\zeta} = \infty$ and define

$$\psi_N = \sum_{k \leq N} \lambda_k^{-1} e_k \langle e_k, \psi\rangle_{L^2_\zeta}, \quad N \in \mathbb{N},$$

which defines an element of $\bar{H}$. By density we can choose $h_N \in H$ such that $\|h_N - \psi_N\|_{L^2_\zeta} < 1/\|\psi\|_{L^2_\zeta}$ as well as $\|I(h_N - \psi_N)\|_{L^2} < 1$, for every N fixed. Next observe that

$$\langle \psi, \psi_N\rangle_{L^2_\zeta} = \sum_{k \leq N} \lambda_k^{-1}\langle e_k, \psi\rangle^2_{L^2_\zeta} \equiv M_N$$

$$\|I(\psi_N)\|^2_{L^2} = \langle I^*I(\psi_N), \psi_N\rangle_{L^2_\zeta} = \sum_{k \leq N} \lambda_k^{-1}\langle e_k, \psi\rangle^2_{L^2_\zeta} = M_N$$

and that $M_N \to \infty$ as $N \to \infty$. Then by our choice of $h_N \in H$ and if $M_N \geq 2$ we have by the triangle inequality,

$$|\langle \psi, h_N\rangle_{L^2_\zeta}| \geq |\langle \psi, \psi_N\rangle_{L^2_\zeta}| - |\langle \psi, \psi_N - h_N\rangle_{L^2_\zeta}| \geq M_N - 1 \geq M_N/2,$$

$$\|I(h_N)\|_{L^2} \leq \|I(\psi_N)\|_{L^2} + \|I(h_N - \psi_N)\|_{L^2} \leq \sqrt{M}_N + 1 \leq 2\sqrt{M}_N.$$

From this we conclude that the inverse of (3.18) satisfies

$$i^{-1}_{\theta,H,\psi} \geq \frac{\langle \psi, h_N\rangle^2_{L^2_\zeta}}{\|Ih_N\|^2_{L^2}} \geq \frac{1}{16}\frac{M_N^2}{M_N} \geq M_N/16.$$

As N was arbitrary and $M_N \to_{N\to\infty} \infty$ we must have $i_{\theta,H,\psi} = 0$, as desired.

□

Index

Bibliography

[1] K. Abraham and R. Nickl. On statistical Caldéron problems. Mathematical Statistics and Learning, (2):165–216, 2019.

[2] R. A. Adams and J. J. F. Fournier. Sobolev spaces. Elsevier/Academic Press, Amsterdam, 2003.

[3] R. J. Adler and J. E. Taylor. Random fields and geometry. Springer Monographs in Mathematics. Springer, New York, 2007.

[4] R. Altmeyer. Polynomial time guarantees for sampling based posterior inference in high-dimensional generalised linear models. manuscript, 2022.

[5] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb. Solving inverse problems using data-driven models. Acta Numer., 28:1–174, 2019.

[6] D. Bakry and M. Émery. Diffusions hypercontractives. In Séminaire de probabilités, XIX, 1983/84, volume 1123 of Lecture Notes in Math., pages 177–206. Springer, Berlin, 1985.

[7] D. Bakry, I. Gentil, and M. Ledoux. Analysis and geometry of Markov diffusion operators, volume 348. Springer, Cham, 2014.

[8] G. Bal and K. Ren. Multi-source quantitative photoacoustic tomography in a diffusive regime. Inverse Problems, 27(7):075003, 20, 2011.

[9] G. Bal and G. Uhlmann. Inverse diffusion theory of photoacoustics. Inverse Problems, 26(8):085010, 20, 2010.

[10] A. S. Bandeira, A. El Alaoui, S. Hopkins, T. Schramm, A. Wein, and I. Zadik. The franz-parisi criterion and computational trade-offs in high dimensional statistics. arXiv, 2022.

[11] A. S. Bandeira, A. Maillard, R. Nickl, and S. Wang. On free energy barriers in gaussian priors and failure of cold start mcmc for high-dimensional unimodal distributions. Phil. Trans. Royal Society, 2023.

[12] A. Barron, M. J. Schervish, and L. Wasserman. The consistency of posterior distributions in nonparametric problems. Ann. Statist., 27(2):536–561, 1999.

[13] R. F. Bass. Diffusions and elliptic operators. Springer, New York, 1998.

[14] A. Belloni and V. Chernozhukov. On the computational complexity of MCMC-based estimators in large samples. Annals of Statistics, 37, 2009.

[15] G. Ben Arous, R. Gheissari, and A. Jagannath. Algorithmic thresholds for tensor PCA. Ann. Probab., 48(4):2052–2087, 2020.

[16] G. Ben Arous, A. Wein, and I. Zadik. Free energy wells and overlap gap property in sparse pca. Comm. Pure Appl. Math., 2022.

[17] M. Benning and M. Burger. Modern regularization methods for inverse problems. Acta Numer., 27:1–111, 2018.

[18] A. Beskos, M. Girolami, S. Lan, P. Farrell, and A. Stuart. Geometric MCMC for infinite-dimensional inverse problems. Journal of Computational Physics, 335:327–351, 2017.

[19] L. Birgé. Approximation dans les espaces métriques et théorie de l'estimation. Z. Wahrsch. Verw. Gebiete, 65(2):181–237, 1983.

[20] L. Birgé. Model selection for gaussian regression with random design. Bernoulli, 10:1039–1051, 2004.

[21] J. Bohr. Stability of the non-abelian X-ray transform in dimension ≥ 3. J. Geom. Anal., 31(11):11226–11269, 2021.

[22] J. Bohr. A Bernstein-von Mises theorem for the Caldéron problem with piece-wise constant conductivities. manuscript, 2022.

[23] J. Bohr and R. Nickl. On log-concave approximations of high-dimensional posterior measures and stability properties in non-linear inverse problems. arXiv, 2021.

[24] A. Bonito, A. Cohen, R. DeVore, G. Petrova, and G. Welper. Diffusion coefficients estimation for elliptic partial differential equations. SIAM J. Math. Anal., 49(2):1570–1592, 2017.

[25] F.-X. Briol, C. J. Oates, M. Girolami, M. A. Osborne, and D. Sejdinovic. Probabilistic integration: a role in statistical computation? Statist. Sci., 34(1):1–22, 2019.

[26] A.-P. Calderón. On an inverse boundary value problem. In Seminar on Numerical Analysis and its Applications to Continuum Physics (Rio de Janeiro, 1980), pages 65–73. Soc. Brasil. Mat., Rio de Janeiro, 1980.

[27] I. Castillo. Pólya tree posterior distributions on densities. Ann. Inst. Henri Poincaré Probab. Stat., 53(4):2074–2102, 2017.

[28] I. Castillo and R. Nickl. Nonparametric Bernstein–von Mises Theorems in Gaussian white noise. Ann. Statist., 41(4):1999–2028, 2013.

[29] I. Castillo and R. Nickl. On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. Ann. Statist., 42(5):1941–1969, 2014.

[30] I. Castillo and J. Rousseau. A Bernstein–von Mises theorem for smooth functionals in semiparametric models. Ann. Statist., 43(6):2353–2383, 2015.

[31] I. Castillo and V. Ročková. Uncertainty quantification for Bayesian CART. Ann. Statist., 49(6):3482–3509, 2021.

[32] I. Castillo and S. van der Pas. Multiscale Bayesian survival analysis. Ann. Statist., 49(6):3559–3582, 2021.

[33] M. F. Chen and S. F. Li. Coupling methods for multidimensional diffusion processes. Ann. Probab., 17(1):151–177, 1989.

[34] K. L. Chung and Z. X. Zhao. From Brownian motion to Schrödinger's equation. Springer-Verlag, Berlin, 1995.

[35] P. Constantin and C. Foias. Navier-Stokes equations. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 1988.

[36] S. Cotter, G. Roberts, A. Stuart, and D. White. MCMC methods for functions: Modifying old algorithms to make them faster. Statistical Science, 28(3):424–446, 2013.

[37] S. L. Cotter, M. Dashti, J. C. Robinson, and A. M. Stuart. Bayesian inverse problems for functions and applications to fluid mechanics. Inverse Problems, 25(11):115008, 43, 2009.

[38] D. D. Cox. An analysis of Bayesian inference for nonparametric regression. Ann. Statist., 21(2):903–923, 1993.

[39] A. S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. J. R. Stat. Soc. Ser. B. Stat. Methodol., 79(3):651–676, 2017.

[40] M. Dashti and A. M. Stuart. The Bayesian approach to inverse problems. In: Handbook of Uncertainty Quantification, Editors R. Ghanem, D. Higdon and H. Owhadi, Springer, 2016.

[41] E. B. Davies. Heat kernels and spectral theory, volume 92 of Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, 1990.

[42] E. B. Davies. Spectral theory and differential operators, volume 42 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 1995.

[43] P. Diaconis. Bayesian Numerical Analysis. In J. Berger and S. Gupta, editors, Statistical Decision Theory and Related Topics IV, pages 163–175. Springer, New York, 1988.

[44] P. Diaconis and D. Freedman. On the consistency of Bayes estimates. Ann. Statist., 14(1):1–67, 1986. With a discussion and a rejoinder by the authors.

[45] S. Dirksen. Tail bounds via generic chaining. Electron. J. Probab., 20:no. 53, 29, 2015.

[46] J. L. Doob. Application of the theory of martingales. In Le Calcul des Probabilités et ses Applications, Colloques Internationaux du Centre National de la Recherche Scientifique, no. 13, pages 23–27. Centre National de la Recherche Scientifique, Paris, 1949.

[47] R. M. Dudley. Real analysis and probability. Cambridge University Press, Cambridge, 2002.

[48] R. M. Dudley. Uniform central limit theorems. Cambridge University Press, New York, second edition, 2014.

[49] A. Durmus and E. Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. Ann. Appl. Probab., 27(3):1551–1587, 2017.

[50] A. Durmus and E. Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. Bernoulli, 25, 2019.

[51] D. E. Edmunds and H. Triebel. Function spaces, entropy numbers, differential operators. Cambridge University Press, Cambridge, 1996.

[52] H. W. Engl, M. Hanke, and A. Neubauer. Regularization of Inverse Problems. Kluwer Academic Publishers Group, 1996.

[53] L. C. Evans. Partial differential equations. American Math. Soc., Second edition, 2010.

[54] D. Freedman. On the Bernstein-von Mises theorem with infinite-dimensional parameters. Ann. Statist., 27(4):1119–1140, 1999.

[55] D. A. Freedman. On the asymptotic behavior of Bayes' estimates in the discrete case. Ann. Math. Statist., 34:1386–1403, 1963.

[56] M. Freidlin. Functional integration and partial differential equations, volume 109 of Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, 1985.

[57] C.-F. Gauß. Theoria Motus Corporum Coelestium. Hamburg, 1809.

[58] S. Ghosal, J. K. Ghosh, and A. W. van der Vaart. Convergence rates of posterior distributions. Ann. Statist., 28(2):500–531, 2000.

[59] S. Ghosal and A. W. van der Vaart. Fundamentals of Nonparametric Bayesian Inference. Cambridge University Press, New York, 2017.

[60] D. Gilbarg and N. S. Trudinger. Elliptic partial differential equations of second order. Springer-Verlag, Berlin-New York, 1998.

[61] E. Giné and R. Nickl. Mathematical foundations of infinite-dimensional statistical models. Cambridge University Press, New York, 2016.

[62] M. Giordano and H. Kekkonen. Bernstein-von mises theorems and uncertainty quantification for linear inverse problems. SIAM /ASA Journal on Uncertainty Quantification, 2020.

[63] M. Giordano and R. Nickl. Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. Inverse Problems, 2020.

[64] M. Giordano and K. Ray. Nonparametric bayesian inference for reversible multi-dimensional diffusions. Ann. Statist., 2022.

[65] S. Gugushvili, A. W. van der Vaart, and D. Yan. Bayesian linear inverse problems in regularity scales. Ann. Inst. Henri Poincaré Probab. Stat., 56(3):2081–2107, 2020.

[66] M. Hairer, A. Stuart, and S. Vollmer. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. The Annals of Applied Probability, 24(6):2455–2490, 2014.

[67] T. Helin and R. Kretschmann. Non-asymptotic error estimates for the laplace approximation in bayesian inverse problems. arXiv preprint 2012.06603, 2020.

[68] A. Hilger, I. Manke, N. Kardjilov, and et al. Tensorial neutron tomography of three-dimensional magnetic vector fields in bulk materials. Nat Commun 9, 4023, 2018.

[69] J. Ilmavirta and F. Monard. Integral geometry on manifolds with boundary and applications. The Radon transform: 100 years, 2019.

[70] K. Ito and K. Kunisch. On the injectivity and linearization of the coefficient-to-solution mapping for elliptic boundary value problems. J. Math. Anal. Appl., 188(3):1040–1066, 1994.

[71] M. Jerrum. Counting, sampling and integrating: algorithms and complexity. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2003.

[72] I. M. Johnstone. High dimensional Bernstein–von Mises: simple examples. In Borrowing strength: theory powering applications—a Festschrift for Lawrence D. Brown, volume 6 of Inst. Math. Stat. (IMS) Collect., pages 87–98. Inst. Math. Statist., Beachwood, OH, 2010.

[73] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. SIAM J. Math. Anal., 29(1):1–17, 1998.

[74] B. Kaltenbacher, A. Neubauer, and O. Scherzer. In Iterative Regularization Methods for Nonlinear Ill-Posed Problems. Radon Series on Computational and Applied Mathematics, 2008.

[75] A. Katchalov, Y. Kurylev, and M. Lassas. Inverse boundary spectral problems. Chapman & Hall/CRC, Boca Raton, FL, 2001.

[76] H. Kekkonen. Consistency of bayesian inference with gaussian process priors for a parabolic inverse problem. Inverse Problems, 2022.

[77] B. Knapik, A. W. van der Vaart, and J. van Zanten. Bayesian inverse problems with gaussian priors. Annals of Statistics, 2011.

[78] J. Kuelbs and W. V. Li. Metric entropy and the small ball problem for Gaussian measures. J. Funct. Anal., 116(1):133–157, 1993.

[79] P.-S. M. d. Laplace. Theorie analytiques des probabilités. Courcier, Paris, 1812.

[80] R. Latał a and D. Matlak. Royen's proof of the Gaussian correlation inequality. In Geometric aspects of functional analysis, volume 2169 of Lecture Notes in Math., pages 265–275. Springer, Cham, 2017.

[81] K. Law, A. Stuart, and K. Zygalakis. Data assimilation, volume 62 of Texts in Applied Mathematics. Springer, Cham, 2015. A mathematical introduction.

[82] L. Le Cam. On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. Univ. California Publ. Statist., 1:277–329, 1953.

[83] L. Le Cam. Asymptotic methods in statistical decision theory. Springer-Verlag, New York, 1986.

[84] H. Leahu. On the Bernstein-von Mises phenomenon in the Gaussian white noise model. Electron. J. Stat., 5:373–404, 2011.

[85] W. V. Li and W. Linde. Approximation, metric entropy and small ball estimates for Gaussian measures. Ann. Probab., 27(3):1556–1578, 1999.

[86] W. V. Li and Q.-M. Shao. Gaussian processes: inequalities, small ball probabilities and applications. In Stochastic processes: theory and methods, volume 19 of Handbook of Statist., pages 533–597. North-Holland, Amsterdam, 2001.

[87] J.-L. Lions and E. Magenes. Non-homogeneous boundary value problems and applications. Vol. I. Springer-Verlag, New York-Heidelberg, 1972.

[88] Y. Lu, A. Stuart, and H. Weber. Gaussian approximations for probability measures on $\mathbb{R}^d$. SIAM/ASA J. Uncertain. Quantif., 5(1):1136–1165, 2017.

[89] F. Monard. Functional relations, sharp mapping properties, and regularization of the X-ray transform on disks of constant curvature. SIAM J. Math. Anal., 52(6):5675–5702, 2020.

[90] F. Monard, R. Nickl, and G. P. Paternain. Efficient nonparametric Bayesian inference for X-ray transforms. Ann. Statist., 47(2):1113–1147, 2019.

[91] F. Monard, R. Nickl, and G. P. Paternain. Consistent inversion of noisy non-Abelian X-ray transforms. Comm. Pure Appl. Math., 74(5):1045–1099, 2021.

[92] F. Monard, R. Nickl, and G. P. Paternain. Statistical guarantees for Bayesian uncertainty quantification in nonlinear inverse problems with Gaussian process priors. Ann. Statist., 49(6):3255–3298, 2021.

[93] A. I. Nachman. Global uniqueness for a two-dimensional inverse boundary value problem. Ann. of Math. (2), 143(1):71–96, 1996.

[94] R. Nickl. Bernstein-von Mises theorems for statistical inverse problems I: Schrödinger equation. J. Eur. Math. Soc., 22: 2697–2750, 2020.

[95] R. Nickl. Inference for diffusions from low frequency measurements. arXiv, 2022.

[96] R. Nickl and G. P. Paternain. On some information-theoretic aspects of non-linear statistical inverse problems. Proc. ICM.,, 2022.

[97] R. Nickl and K. Ray. Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. Ann. Statist., 48(3):1383–1408, 2020.

[98] R. Nickl and J. Söhl. Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. Ann. Statist., 45(4):1664–1693, 2017.

[99] R. Nickl and J. Söhl. Bernstein-von Mises theorems for statistical inverse problems II: compound Poisson processes. Electron. J. Stat., 13(2):3513–3571, 2019.

[100] R. Nickl and E. Titi. Posterior consistency of bayesian data assimilation with gaussian process priors for the 2D navier stokes equations. manuscript, 2023.

[101] R. Nickl, S. van de Geer, and S. Wang. Convergence rates for penalised least squares estimators in PDE-constrained regression problems. SIAM J. Uncert. Quant., 8, 2020.

[102] R. Nickl and S. Wang. On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms. J. Eur. Math. Soc., 2022.

[103] G. P. Paternain and M. Salo. The non-abelian x-ray transform on surfaces. Journal of Diff. Geometry, 2022, to appear.

[104] G. P. Paternain, M. Salo, and G. Uhlmann. The attenuated ray transform for connections and Higgs fields. Geom. Funct. Anal., 22(5):1460–1489, 2012.

[105] G. P. Paternain, M. Salo, and G. Uhlmann. Geometric inverse problems in 2D. Cambridge University Press, Cambridge, UK, 2023.

[106] L. Pestov and G. Uhlmann. Two dimensional compact simple Riemannian manifolds are boundary distance rigid. Ann. of Math. (2), 161(2):1093–1110, 2005.

[107] J. Radon. Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. Berichte über die Verhandlungen der Königlich-Sächsischen Akademie der Wissenschaften zu Leipzig. Mathematisch-Physische Klasse 69 262–277, 1917.

[108] K. Ray. Bayesian inverse problems with non-conjugate priors. Electron. J. Stat., 7:2516–2549, 2013.

[109] S. Reich and C. Cotter. Probabilistic forecasting and Bayesian data assimilation. Cambridge University Press, New York, 2015.

[110] M. Reiß. Asymptotic equivalence for nonparametric regression with multivariate and random design. Ann. Statist., 36(4):1957–1982, 2008.

[111] G. R. Richter. An inverse problem for the steady state diffusion equation. SIAM J. Appl. Math., 41(2):210–221, 1981.

[112] V. Rivoirard and J. Rousseau. Bernstein-von Mises theorem for linear functionals of the density. Ann. Statist., 40(3):1489–1523, 2012.

[113] C. P. Robert and G. Casella. Monte Carlo statistical methods. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2004.

[114] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. Bernoulli, 2(4):341–363, 1996.

[115] M. Sales, M. Strobl, T. Shinohara, and et al. Three dimensional polarimetric neutron tomography of magnetic fields. Sci Rep 8, 2214, 2018.

[116] C. Schillings, B. Sprungk, and P. Wacker. On the convergence of the Laplace approximation and noise-level-robustness of Laplace-based Monte Carlo methods for Bayesian inverse problems. Numer. Math., 145(4):915–971, 2020.

[117] L. Schwartz. On Bayes procedures. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 4:10–26, 1965.

[118] V. Spokoiny. Dimension free non-asymptotic bounds on the accuracy of high dimensional laplace approximation. arXiv preprint 2204.11038, 2022.

[119] A. M. Stuart. Inverse problems: a Bayesian perspective. Acta Numer., 19:451–559, 2010.

[120] J. Sylvester and G. Uhlmann. A global uniqueness theorem for an inverse boundary value problem. Ann. of Math. (2), 125(1):153–169, 1987.

[121] M. Talagrand. Upper and lower bounds for stochastic processes. Springer, Heidelberg, 2014.

[122] M. E. Taylor. Partial differential equations I. Basic theory. Springer, New York, 2011.

[123] M. E. Taylor. Partial differential equations II. Qualitative studies of linear equations. Springer, New York, 2011.

[124] G. Uhlmann. Electrical impedance tomography and Calderón's problem. Inverse Problems, 25(12):123011, 39, 2009.

[125] S. A. van de Geer. Applications of empirical process theory. Cambridge University Press, Cambridge, 2000.

[126] A. W. van der Vaart. On differentiable functionals. Ann. Statist., 19(1):178–204, 1991.

[127] A. W. van der Vaart. Asymptotic statistics. Cambridge Univ. Press, 1998.

[128] A. W. van der Vaart and J. H. van Zanten. Bayesian inference with rescaled Gaussian process priors. Electron. J. Stat., 1:433–448, 2007.

[129] A. W. van der Vaart and J. H. van Zanten. Rates of contraction of posterior distributions based on Gaussian process priors. Ann. Statist., 36(3):1435–1463, 2008.

[130] A. W. van der Vaart and J. A. Wellner. Weak convergence and empirical processes. Springer Series in Statistics. Springer-Verlag, New York, 1996.

[131] J. van Neerven. γ-radonifying operators—a survey. volume 44 of Proc. Centre Math. Appl. Austral. Nat. Univ., pages 1–61. Austral. Nat. Univ., Canberra, 2010.

[132] R. Vershynin. High-dimensional probability. Cambridge University Press, Cambridge, 2018.

[133] C. Villani. Optimal transport. Old and new. Springer-Verlag, Berlin, 2009.

[134] S. J. Vollmer. Posterior consistency for Bayesian inverse problems through stability and regression results. Inverse Problems, 29(12):125011, 32, 2013.

[135] R. Wong. Asymptotic approximations of integrals. SIAM, 2001. Corrected reprint of the 1989 original.