

Consistency of Bayesian inference with Gaussian process priors for a parabolic inverse problem

Kekkonen, Hanne

DOI

[10.1088/1361-6420/ac4839](https://doi.org/10.1088/1361-6420/ac4839)

Publication date

2022

Document Version

Final published version

Published in

Inverse Problems

Citation (APA)

Kekkonen, H. (2022). Consistency of Bayesian inference with Gaussian process priors for a parabolic inverse problem. *Inverse Problems*, 38(3), Article 035002. <https://doi.org/10.1088/1361-6420/ac4839>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

PAPER • OPEN ACCESS

Consistency of Bayesian inference with Gaussian process priors for a parabolic inverse problem

To cite this article: Hanne Kekkonen 2022 *Inverse Problems* **38** 035002

View the [article online](#) for updates and enhancements.

You may also like

- [Tenth International Conference on Topics in Astroparticle and Underground Physics \(TAUP2007\)](#)
Kunio Inoue, Atsuto Suzuki and Tadao Mitsui
- [Semiclassical theory of resonant dissociative excitation of molecular ions by electron impact](#)
A A Narits, K S Kislov and V S Lebedev
- [Global solvability and stabilization to a cancer invasion model with remodelling of ECM](#)
Chunhua Jin



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Consistency of Bayesian inference with Gaussian process priors for a parabolic inverse problem

Hanne Kekkonen* 

Delft Institute of Applied Mathematics, Delft University of Technology, The Netherlands

E-mail: h.n.kekkonen@tudelft.nl

Received 9 September 2021, revised 21 December 2021

Accepted for publication 5 January 2022

Published 25 January 2022



Abstract

We consider the statistical non-linear inverse problem of recovering the absorption term $f > 0$ in the heat equation

$$\begin{cases} \partial_t u - \frac{1}{2} \Delta u + fu = 0 & \text{on } \mathcal{O} \times (0, \mathbf{T}) \\ u = g & \text{on } \partial \mathcal{O} \times (0, \mathbf{T}) \\ u(\cdot, 0) = u_0 & \text{on } \mathcal{O}, \end{cases}$$

where $\mathcal{O} \in \mathbb{R}^d$ is a bounded domain, $\mathbf{T} < \infty$ is a fixed time, and g, u_0 are given sufficiently smooth functions describing boundary and initial values respectively. The data consists of N discrete noisy point evaluations of the solution u_f on $\mathcal{O} \times (0, \mathbf{T})$. We study the statistical performance of Bayesian nonparametric procedures based on a large class of Gaussian process priors. We show that, as the number of measurements increases, the resulting posterior distributions concentrate around the true parameter generating the data, and derive a convergence rate for the reconstruction error of the associated posterior means. We also consider the optimality of the contraction rates and prove a lower bound for the minimax convergence rate for inferring f from the data, and show that optimal rates can be achieved with truncated Gaussian priors.

Keywords: nonlinear inverse problems, parabolic PDEs, Bayesian inference, Gaussian priors, frequentist consistency

*Author to whom any correspondence should be addressed.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

1. Introduction

Inverse problems arise from the need to extract information from indirect and noisy measurements. In many scientific disciplines, such as imaging, medicine, material sciences and engineering, the relationship between the quantity of interest and the collected data is determined by the physics of the underlying system and can be modelled mathematically. In general, we are interested in recovering some function f from measurements of $G(f)$, where G is the forward operator of some partial differential equation (PDE). In practice, a statistical observation scheme provides us data

$$Y_i = G(f)(Z_i) + \sigma W_i, \quad i = 1, \dots, N, \quad (1)$$

where the Z_i 's are points at which the PDE solution $G(f)$ is measured, and the W_i 's are standard Gaussian noise variables scaled by a fixed noise level $\sigma > 0$. The inverse problem then consists of reconstructing f from the noisy measurements $(Y_i, Z_i)_{i=1}^N$. Many, possibly nonlinear, inverse problems fit into this framework, including electrical impedance tomography [15, 31], photoacoustic tomography and several other hybrid imaging problems [8, 9, 39], and inverse scattering [16, 30]. Even though inverse problems have been studied in great detail, see e.g. [11, 21, 33], statistical noise models, such as the one above, have been analysed only more recently [6, 13, 32].

In many applications the forward operator G arising from the related PDE is non-linear in f , and so the negative log-likelihood function arising from the measurement (1) can be non-convex. This means that many commonly used methods like Tikhonov regularisation and maximum *a priori* (MAP) estimation, where one has to minimise a penalised log-likelihood function, cannot be reliably computed by standard convex optimisation techniques. There are some iterative optimisation methods, such as Landweber iteration and Levenberg–Marquardt regularisation, that circumvent the problems arising from non-convexity, see [34] where the parabolic PDE considered here has been studied, and also [7, 10, 33]. In this paper we consider the Bayesian approach which offers an attractive alternative for solving complex inverse problems, see e.g. [18, 57]. In the standard Bayesian approach one assigns a Gaussian prior Π to f , which is then updated, given data $(Y_i, Z_i)_{i=1}^N$, into a posterior distribution for f , using Bayes' theorem. The posterior distribution can be used to calculate point estimates but it also delivers an estimate of the statistical uncertainty in the reconstruction. If the forward map can be evaluated numerically one can deploy modern MCMC methods, such as stochastic gradient MCMC and parallel tempering, to construct computationally efficient Bayesian algorithms even for complicated non-linear inverse problems [12, 20, 40, 44], hence avoiding optimisation algorithms and inversion of G . Computational guarantees for the mixing times of such algorithms are also available even in general high-dimensional non-linear settings [14, 28, 53].

Since there is no objective way to select a prior distribution it is natural to ask how the choice of the prior affects the solution, and especially if the conclusions are asymptotically independent of the prior. Another important question that arises is whether Bayesian inference provides a statistically optimal estimate of the unknown quantity f . If we assume that the data are generated from a fixed 'true' function $f = f_0$, we would like to know whether the posterior mean $\bar{f} = \mathbb{E}^\Pi(f | (Y_i, Z_i)_{i=1}^N)$ converges towards the ground truth, and at what speed the posterior contracts around f_0 . Nonparametric Bayesian inverse problems have been extensively studied in linear settings and the statistical validity of Bayesian inversion methods is quite well understood, see e.g. [2, 26, 35–37, 46, 55, 58].

However, non-linear inverse problems are fundamentally more challenging and very little is known about the frequentist performance of Bayesian methods. Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions were considered in [50], and Bernstein–von Mises theorems for time-independent Schrödinger equation and compound Poisson process were studied in [49, 51] respectively. The frequentist consistency of Bayesian inversion in the elliptic PDE in divergence form was examined in [61]. All the above papers employ ‘uniform wavelet type priors’ with bounded C^β -norms. In this paper we consider more practical Gaussian process priors which are regularly used in applications, and which allow the use of modern MCMC methods such as the preconditioned Crank–Nicholson algorithm. Gaussian process priors were considered in [27] where the consistency of the Bayesian approach in the nonlinear inverse problem of reconstructing the diffusion coefficient from noisy observations of the solution to an elliptic PDE in divergence form was studied. Notably, building on the ideas from [48] where the consistency of Bayesian inversion of noisy non-abelian x-ray transform is considered [27], also provides contraction results for general non-linear inverse problems that fulfil certain Lipschitz and stability conditions. A general class of non-linear inverse regression models, satisfying particular analytic conditions on the model including invertibility of the related Fisher information operator, has been considered in the recent paper [47], where a general semi-parametric Bernstein–von Mises theorem is proved. Closely related to [27] are the results achieved in [52] for MAP estimates associated to Gaussian process priors, but since the proofs are based on variational methods they are very different from the Bayesian ones. We also mention the recent results on statistical Caldéron problem [1], where a logarithmic contraction speed is proved for the problem of recovering an unknown conductivity function from noisy measurements of the voltage to current map, also known as the Dirichlet-to-Neumann map, at the boundary of the medium.

In this paper we consider the problem of recovering a coefficient of a parabolic partial differential operator from observations of a solution to the associated PDE, under given boundary and initial value conditions, corrupted by additive Gaussian noise. More precisely, we will study the heat equation with an additional absorption or cooling term that presents all the conceptual difficulties of a time dependent parabolic PDEs but allows a clean exposition; let g and u_0 be sufficiently smooth boundary and initial value functions respectively, and let $f : \mathcal{O} \rightarrow \mathbb{R}$ be an unknown absorption term determining the solutions u_f of the PDE

$$\begin{cases} L_f(u) = \partial_t u - \frac{1}{2} \Delta_x u + f u = 0 & \text{on } \mathcal{O} \times (0, \mathbf{T}) \\ u = g & \text{on } \partial \mathcal{O} \times (0, \mathbf{T}) \\ u(\cdot, 0) = u_0 & \text{on } \mathcal{O}, \end{cases} \quad (2)$$

where $\Delta_x = \sum_{i=1}^d \partial^2 / \partial x_i^2$ denotes the standard Laplace operator and ∂_t the time derivative. Under mild regularity conditions on f , assuming that $f > 0$, and g, u_0 satisfying natural consistency conditions on $\partial \mathcal{O} \times \{0\}$ the theory of parabolic PDEs implies that (2) has a unique classical solution $G(f) = u_f \in C(\overline{\mathcal{O}} \times \overline{(0, \mathbf{T})}) \cap C^{2,1}(\mathcal{O} \times (0, \mathbf{T}))$. The above type reaction–diffusion equations can also be used to describe ecological systems like population dynamics, with u being density of prey and f describing resources or the effect of predators [62], evolution of competing languages [54, 56], and many other spread phenomena. Another attractive way of modelling time evolution is using stochastic PDEs, see e.g. [3, 29] and the references therein. However, these models are usually not feasible for Bayesian analysis due to complex likelihood functions.

We show that the posterior means arising from a large class of Gaussian process priors for f provide statistically consistent recovery of the unknown function f in (2) given data (1), where Z_i are drawn from a uniform distribution on $\mathcal{O} \times (0, \mathbf{T})$, and provide explicit polynomial convergence rates as the number of measurements increases. We start with contraction results for posterior distributions arising from a wide class of rescaled Gaussian process priors, similar to those considered in [27, 48], which address the need for additional *a priori* regularisation of the posterior distribution to overcome the effects of non-linearity of the forward map G . Building on the ideas from [48] and further developed in [27], we first show that the posterior distributions arising from these priors optimally solve the PDE-constrained regression problem of inferring $G(f)$ from the data (1). These results can then be combined with suitable stability estimates for the inverse of G to show that the posterior distribution contracts around the true parameter f_0 , that generated the data, at certain polynomial rate, and that the posterior mean converges to the truth with the same rate. We also consider the optimality of the contraction rates and prove a lower bound for the minimax convergence rate for inferring f_0 from the data. We note that, while the rates achieved in the first part approach the optimal rate for very smooth models, they are not in general optimal. In the second part of the paper we show that optimal rates can be achieved with truncated and rescaled Gaussian priors. To the best of our knowledge this is the first time such optimality results are shown for a non-linear inverse problem.

This paper is organised as follows. The basic setting of the statistical inverse problem and the notations used in the paper can be found in section 2. The main results are stated in section 3 and their proofs are given in section 4.

2. A statistical inverse problem for parabolic PDEs

2.1. Parabolic Hölder and Sobolev spaces

Throughout this paper $\mathcal{O} \in \mathbb{R}^d$, $d \in \mathbb{N}$, is a given non-empty, open and bounded set with smooth boundary $\partial\mathcal{O}$ and closure $\overline{\mathcal{O}}$. We define the space-time cylinder $\mathcal{Q} = \mathcal{O} \times (0, \mathbf{T})$, with $\mathbf{T} \in (0, \infty)$, and denote its lateral boundary $\partial\mathcal{O} \times (0, \mathbf{T})$ by Σ .

The spaces of continuous functions defined on $\mathcal{O} \subset \mathbb{R}^d$ and $\overline{\mathcal{O}}$ are denoted by $C(\mathcal{O})$ and $C(\overline{\mathcal{O}})$ respectively, and endowed with the supremum norm $\|\cdot\|_\infty$. For positive integers $k \in \mathbb{N}$, $C^k(\mathcal{O})$ is the space of k -times differentiable functions with uniformly continuous derivatives. For non-integer $s > 0$ we define $C^s(\mathcal{O})$ as

$$C^s(\mathcal{O}) = \left\{ f \in C^{\lfloor s \rfloor}(\mathcal{O}) : \forall |\alpha| = \lfloor s \rfloor, \sup_{x,y \in \mathcal{O}, x \neq y} \frac{|D^\alpha f(x) - D^\alpha f(y)|}{|x - y|^{s - \lfloor s \rfloor}} < \infty \right\},$$

where $\lfloor s \rfloor$ denotes the largest integer less than or equal to s , and for any multi-index $\alpha = (\alpha_1, \dots, \alpha_d)$, D^α is the α th partial differential operator. The Hölder space $C^s(\mathcal{O})$ is normed by

$$\|f\|_{C^s(\mathcal{O})} = \sum_{|\alpha| \leq \lfloor s \rfloor} \|D^\alpha f(x)\|_\infty + \sum_{|\alpha| = \lfloor s \rfloor} \sup_{x,y \in \mathcal{O}, x \neq y} \frac{|D^\alpha f(x) - D^\alpha f(y)|}{|x - y|^{s - \lfloor s \rfloor}},$$

where the second sum is removed if s is an integer. We denote by $C^\infty(\mathcal{O}) = \bigcap_s C^s(\mathcal{O})$ the set of smooth functions. We also need Hölder–Zygmund spaces $\mathcal{C}^s(\mathcal{O})$ which can be defined as a special case of Besov spaces by $\mathcal{C}^s(\mathcal{O}) = B_{\infty, \infty}^s(\mathcal{O})$, $s \geq 0$, see [59, section 3.4.2] for definitions. If $s \notin \mathbb{N}$ then $\mathcal{C}^s(\mathcal{O}) = C^s(\mathcal{O})$ and we have the continuous embeddings $\mathcal{C}^{s'}(\mathcal{O}) \subsetneq C^s(\mathcal{O}) \subsetneq \mathcal{C}^s(\mathcal{O})$, for $s \in \mathbb{N} \cup \{0\}$, $s' > s$.

The classical space to look for a solution to (2) is the parabolic Hölder space $C^{2,1}(Q)$ defined by

$$C^{2,1}(Q) = \{f \in C(Q) : \exists \partial_t f, D_{ij}f \in C(Q), \quad i, j = 1, \dots, d\}.$$

Let $\theta \in (0, 1]$ and define

$$\|f\|_{C^{\theta,\theta/2}(Q)} = \|f\|_{\infty} + [f]_{\theta,\theta/2}, \quad [f]_{\theta,\theta/2} = \sup_{z_1, z_2 \in Q, z_1 \neq z_2} \frac{|f(z_1) - f(z_2)|}{\rho(z_1, z_2)^{\theta}},$$

where $\rho(z_1, z_2)$ is the parabolic distance between points $z_1 = (x_1, t_1) \in \mathbb{R}^{d+1}$ and $z_2 = (x_2, t_2) \in \mathbb{R}^{d+1}$ given by $\rho(z_1, z_2) = (\|x_1 - x_2\|_2^2 + |t_1 - t_2|)^{1/2}$. We denote by $C^{\theta,\theta/2}(Q)$ the space of all functions f for which $\|f\|_{C^{\theta,\theta/2}} < \infty$. Finally, the parabolic Hölder space $C^{2+\theta,1+\theta/2}(Q)$, $\theta \in (0, 1]$ is defined as the space of all functions f for which

$$\|f\|_{C^{2+\theta,1+\theta/2}(Q)} = \sum_{|\alpha| \leq 2} \|D^{\alpha} f(x, t)\|_{\infty} + \|\partial_t f\|_{\infty} + [f]_{2+\theta,1+\theta/2} < \infty,$$

where

$$[f]_{2+\theta,1+\theta/2} = \sum_{i,j=1}^d [D_{ij}f]_{\theta,\theta/2} + [\partial_t f]_{\theta,\theta/2}.$$

Parabolic Hölder spaces are Banach spaces. For further details see e.g. [5, 23, 38]. Higher order parabolic Hölder spaces can be defined in a similar way. We will also need parabolic Hölder–Zygmund (Besov–Hölder) spaces $C^{2+\theta,1+\theta/2}(Q) = B_{\infty,\infty}^{2+\theta,1+\theta/2}(Q)$, which possess similar properties to the isotropic Hölder–Zygmund spaces, see e.g. [4, chapter 7.2].

Denote by $L^2(\mathcal{O})$ the Hilbert space of square integrable functions on \mathcal{O} , equipped with its usual inner product $\langle \cdot, \cdot \rangle_{L^2(\mathcal{O})}$. For an integer $k \geq 0$, the order- k Sobolev space on \mathcal{O} is the separable Hilbert space

$$H^k(\mathcal{O}) = \{f \in L^2(\mathcal{O}) : \forall |\alpha| \leq k, \exists D^{\alpha} f \in L^2(\mathcal{O})\},$$

with the inner product $\langle f, g \rangle_{H^k(\mathcal{O})} = \sum_{|\alpha| \leq k} \langle D^{\alpha} f, D^{\alpha} g \rangle_{L^2(\mathcal{O})}$. For a non-integer $s \geq 0$, $H^s(\mathcal{O})$ can be defined by interpolation, see, e.g. [41, section 1.9.1].

We will also use parabolic Sobolev spaces $H^{s,s/2}(Q)$, with $s \geq 0$, defined by

$$H^{s,s/2}(Q) = L^2((0, \mathbf{T}); H^s(\mathcal{O})) \cap H^{s/2}((0, \mathbf{T}); L^2(\mathcal{O})),$$

which is a Hilbert space with a norm

$$\|u\|_{H^{s,s/2}(Q)}^2 = \int_0^{\mathbf{T}} \|u(\cdot, t)\|_{H^s(\mathcal{O})}^2 dt + \|u\|_{H^{s/2}((0, \mathbf{T}); L^2(\mathcal{O}))}^2,$$

see [42, section 4.2.1].

For $s > d/2$ the Sobolev embedding theorem implies that $H^s(\mathcal{O})$ embeds continuously into $C^r(\mathcal{O})$ for any $s > r + d/2 \geq d/2$. Let $g \in H^{s,s/2}(Q)$ and $f \in H^s(\mathcal{O})$. We then have

$$\|fg\|_{H^{s,s/2}(Q)} \leq c \|f\|_{H^s(\mathcal{O})} \|g\|_{H^{s,s/2}(Q)}, \quad s > d/2. \quad (3)$$

We will also use the following bound for $g \in H^{s,s/2}(\mathcal{Q})$ and $f \in C^{s,s/2}(\mathcal{Q})$

$$\|fg\|_{H^{s,s/2}(\mathcal{Q})} \leq c \|f\|_{C^{s,s/2}(\mathcal{Q})} \|g\|_{H^{s,s/2}(\mathcal{Q})}, \quad s \geq 0. \quad (4)$$

The above bounds follow from similar results in isotropic spaces (see e.g. [59]).

Whenever there is no risk of confusion, we will omit the reference to the underlying domain \mathcal{O} or \mathcal{Q} . Attaching a subscript c to any space X denotes the subspace $(X_c, \|\cdot\|_X)$ consisting of functions with compact support in \mathcal{O} (or \mathcal{Q}). Also, if K is a non-empty compact subset of \mathcal{O} , we denote by $H_K^s(\mathcal{O})$ the closed subspace of functions in $H^s(\mathcal{O})$ with support contained in K . The above definitions extend without difficulty to the case where \mathcal{Q} is replaced by its lateral boundary Σ .

We use the symbols \lesssim and \gtrsim for inequalities holding up to a universal constant. For a sequence of random variables F_N we write $F_N = O_P(a_N)$ if for all $\varepsilon > 0$ there exists $M_\varepsilon < \infty$ such that $P(|F_N| \geq M_\varepsilon a_N) < \varepsilon$ for all N large enough. Finally, we denote by $\mathcal{L}(\mathcal{F})$ the law of a random variable F .

2.2. The measurement model and Bayesian approach

Let the source function $g \in C^{2+\theta,1+\theta/2}(\overline{\Sigma})$ and the initial value function $u_0 \in C^{2+\theta}(\overline{\mathcal{O}})$, with $\theta \in (0, 1]$ satisfy the consistency conditions

$$g(x, 0) = u_0(x) \quad \text{and} \quad \partial_t g(x, 0) - \frac{1}{2} \Delta_x u_0(x) + f(x)u_0(x) = 0, \quad x \in \partial\mathcal{O}. \quad (5)$$

Also, let $f \in C^\theta(\mathcal{O})$ and $f \geq f_{\min} > 0$. Then the initial boundary value problem (2) has a unique classical solution $u_f \in C(\overline{\mathcal{O}}) \cap C^{2+\theta,1+\theta/2}(\mathcal{Q})$, see e.g. [38, 43], and a representation in terms of the Feynman–Kac formula

$$\begin{aligned} u(x, t) = & \mathbb{E}^x \left(u_0(X_t) 1_{\{\tau_t=t\}} \exp \left(- \int_0^t f(X_s) ds \right) \right) \\ & + \mathbb{E}^x \left(g(X_{\tau_{\mathcal{O}}}, \tau_{\mathcal{O}}) 1_{\{\tau_t < t\}} \exp \left(- \int_0^{\tau_{\mathcal{O}}} f(X_s) ds \right) \right) \quad (x, t) \in \mathcal{Q}. \quad (6) \end{aligned}$$

Above 1_A is the indicator function of a subset A , $(X_s : s \geq 0)$ is a d -dimensional Brownian motion started at $x \in \mathcal{O}$, with the exit time $\tau_{\mathcal{O}}$ satisfying $\sup_{x \in \mathcal{O}} \mathbb{E}^x(\tau_{\mathcal{O}}) < \infty$, and $\tau_t = \min\{\tau_{\mathcal{O}}, t\}$, see e.g. [22]. The related inverse problem is to recover f given u_f (and g, u_0). If we additionally assume that f is bounded, $g \geq g_{\min} > 0$ and $u_0 \geq u_{0,\min} > 0$ we see, using (6) and Jensen's inequality, that $u_f > 0$. Hence, given u_f we can simply write $f = \frac{(\frac{1}{2}\Delta_x - \partial_t)u_f}{u_f}$. The more practical question we are interested in, is how to optimally solve the above non-linear inverse problem when the observations are corrupted by statistical noise.

We consider the following parameter space for f : for an integer $\alpha > 2 + d/2, f_{\min} > 0$, and $n = n(x)$ being the outward pointing normal at $x \in \partial\mathcal{O}$, let

$$\mathcal{F}_{\alpha, f_{\min}} = \left\{ f \in H^\alpha(\mathcal{O}) : \inf_{x \in \mathcal{O}} f(x) > f_{\min}, f|_{\partial\mathcal{O}} = 1, \frac{\partial^j f}{\partial n^j}|_{\partial\mathcal{O}} = 0 \quad \text{for } 1 \leq j \leq \alpha - 1 \right\}. \quad (7)$$

Let $f \in \mathcal{F}_{\alpha, f_{\min}}$ and denote by $G(f)$ the solution to (2). The measurement model we consider is

$$Y_i = G(f)(Z_i) + \sigma W_i, \quad W_i \sim \mathcal{N}(0, 1) \quad i = 1, \dots, N, \quad (8)$$

where the noise amplitude σ is considered to be a known constant, and the design points $Z_i = (X_i, T_i)$ are drawn from a uniform distribution on \mathcal{Q} . That is, for $N \in \mathbb{N}$,

$$(Z_i)_{i=1}^N \sim \mu, \quad \mu = dz/\text{vol}(\mathcal{Q}),$$

with dz being the Lebesgue measure and $\text{vol}(\mathcal{Q}) = \int_{\mathcal{Q}} d(x, t)$. We assume that both space and time variable follow uniform distribution to unify the following approach. The results could also be developed for deterministic time design at the expense of introducing further technicalities.

We will take the Bayesian approach to the inverse problem of inferring f from the noisy measurements $(Y_i, Z_i)_{i=1}^N$, and place *a priori* measure on the unknown parameter f . Gaussian process priors are a natural choice but they are supported in linear spaces (e.g., $H_c^\alpha(\mathcal{O})$), which is why we next define a convenient bijective re-parametrisation for $f \in \mathcal{F}_{\alpha, f_{\min}}$. We follow the approach of using regular link functions as in [27, 52].

Definition 1.

- (a) A function Φ is called a link function if it is a smooth, strictly increasing bijective map $\Phi: \mathbb{R} \rightarrow (f_{\min}, \infty)$ satisfying $\Phi(0) = 1$ and $\Phi'(t) > 0$ for all $t \in \mathbb{R}$.
- (b) A function $\Phi: (a, b) \rightarrow \mathbb{R}$, $-\infty \leq a, b \leq \infty$, is called regular if all the derivatives of Φ are bounded on \mathbb{R} .

Note that given any link function Φ , one can show (see [52, section 3.1]) that the parameter space $\mathcal{F}_{\alpha, f_{\min}}$ in (7) can be written as

$$\mathcal{F}_{\alpha, f_{\min}} = \{\Phi \circ F : F \in H_c^\alpha(\mathcal{O})\}.$$

We can then consider the solution map associated to (2) as one defined on $H_c^\alpha(\mathcal{O})$;

$$\mathcal{G}: H_c^\alpha(\mathcal{O}) \rightarrow L^2(\mathcal{O}), \quad F \mapsto \mathcal{G}(F) := G(\Phi \circ F), \quad (9)$$

where $G(\Phi \circ F) = G(f)$ is the solution to (2) with $f = \Phi \circ F \in \mathcal{F}_{\alpha, f_{\min}}$.

Using the above re-parametrisation $f = \Phi \circ F$ with a given link function the observation scheme (8) can be rewritten as

$$Y_i = \mathcal{G}(F)(Z_i) + \sigma W_i, \quad i = 1, \dots, N. \quad (10)$$

The random vectors (Y_i, Z_i) on $\mathbb{R} \times \mathcal{Q}$ are then i.i.d. with laws denoted by P_F^i . It follows that P_F^i has the Radon–Nikodym density

$$p_F(y, z) := \frac{dP_F^i}{dy \times d\mu}(y, z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y - \mathcal{G}(F)(z))^2}{2\sigma^2}}, \quad y \in \mathbb{R}, \quad z \in \mathcal{Q}, \quad (11)$$

where dy denotes the Lebesgue measure on \mathbb{R} . We write $P_F^N = \otimes_{i=1}^N P_F^i$ for the joint law of $(Y_i, Z_i)_{i=1}^N$ on $\mathbb{R}^N \times \mathcal{Q}^N$, and $\mathbb{E}_F^i, \mathbb{E}_F^N$ for the expectation operators corresponding to the laws P_F^i, P_F^N respectively.

We model the parameter $F \in H_c^\alpha(\mathcal{O})$ by a Borel probability measure Π supported on the Banach space $C(\mathcal{O})$. Since the map $(F, (y, z)) \mapsto p_F(y, z)$ can be shown to be jointly measurable the posterior distribution $\Pi(\cdot|Y^N, Z^N)$ of $F|Y^N, Z^N$ arising from the model (10) equals to

$$\Pi(B|Y^N, Z^N) = \frac{\int_B e^{\ell^N(F)} d\Pi(F)}{\int_{C(\mathcal{O})} e^{\ell^N(F)} d\Pi(F)}$$

for any Borel set $B \subseteq C(\mathcal{O})$. Above

$$\ell^N(F) = -\frac{1}{2\sigma^2} \sum_{i=1}^N (Y_i - \mathcal{G}(F)(Z_i))^2$$

is the joint log-likelihood function up to a constant.

3. Posterior consistency results

We consider priors that are build around a Gaussian base prior Π' and then rescaled by $N^{-\gamma}$, with appropriate $\gamma > 0$, to provide additional regularisation to combat the non-linearity of the inverse problem, as suggested in [48] and further studied in [27], see also [52]. We first consider certain Gaussian process priors supported on C^β , $\beta \geq 2$, and show that the posterior distributions arising from these priors concentrate near sufficiently regular ground truth F_0 (or f_0) assuming that the data (Y^N, Z^N) is generated through model (10) with $F = F_0$. We then prove a minimax lower bound for inferring f from the data, and show that the optimal convergence rate can be achieved using truncated Gaussian base priors. The proofs of the theorems can be found from section 4.

In the following we are interested in recovering F_0 (or f_0) with H^α , $\alpha > \beta + d/2 \geq 2 + d/2$, smoothness. For this we assume that $g \in H^{3/2+\alpha, 3/4+\alpha/2}(\Sigma)$ and $u_0 \in H^{1+\alpha}(\mathcal{O})$ satisfy the following consistency condition; there exists $\psi \in H^{2+\alpha, 1+\alpha/2}(\bar{Q})$ such that

$$\psi(x, t) = g(x, t) \quad \text{on } \Sigma, \quad \psi(x, 0) = u_0(x) \quad \text{on } \mathcal{O} \quad (12)$$

and ψ satisfies

$$\partial_t^k \left(\left(\partial_t - \frac{1}{2} \Delta_x + f \right) \psi \right) \Big|_{t=0} = 0 \quad \text{for } 0 \leq k < \frac{\alpha}{2} - \frac{1}{2}. \quad (13)$$

Then $u_f \in H^{2+\alpha, 1+\alpha/2}(\bar{Q}) \subset C^{2+\beta, 1+\beta/2}(\bar{Q})$, [42, theorem 5.3]. The above is a standard compatibility condition and a generalisation of (5) stating that the source and initial value functions meet smoothly enough on the boundary $\partial\mathcal{O} \times \{0\}$ see, e.g., [42, section 2] for general compatibility relations and trace theorems in parabolic spaces. Especially we have that

$$g(x, 0) = u_0(x), \quad x \in \partial\mathcal{O}.$$

Note that if $f \in \mathcal{F}_{\alpha, f_{\min}}$ then $f|_{\partial\mathcal{O}} = 1$ and we can, for example, take $g(x, t) = 1 - t$ and $u_0(x) = 1$ in a small neighbourhood of $\partial\mathcal{O} \times \{0\}$. The above consistency conditions are sufficient but can be relaxed for many of the following results. We also assume that

$$u_0 \geq u_{0, \min} > 0, \quad \text{and} \quad g \geq g_{\min} > 0 \quad (14)$$

so that $u_f > 0$ for $f \in \mathcal{F}_{\alpha, f_{\min}}$.

3.1. Rescaled Gaussian priors

We refer, e.g., to [25, section 2] for the basic definitions of Gaussian measures and their reproducing kernel Hilbert spaces (RKHS).

Condition 2. Let $\alpha > \beta + d/2$, with some $\beta \geq 2$, and let \mathcal{H} be a Hilbert space continuously imbedded into $H_c^\alpha(\mathcal{O})$. Let Π' be a centred Gaussian Borel probability measure on the Banach space $C(\mathcal{O})$ that is supported on a separable measurable linear subspace of $C^\beta(\mathcal{O})$, and assume that the reproducing-kernel Hilbert space of Π' equals to \mathcal{H} .

As a simple example of a base prior satisfying condition 2 we can consider Whittle–Matérn process $M = \{M(x), x \in \mathcal{O}\}$ of regularity $\alpha - d/2$, see [27, example 25] or [24, example 11.8] for details. Assume that the measurement (10) is generated from a ‘true unknown’ $F_0 \in H^\alpha(\mathcal{O})$ that is supported on a given compact subset K of the domain \mathcal{O} , and fix a smooth cutoff function $\chi \in C_c^\infty(\mathcal{O})$ such that $\chi = 1$ on K . Then $\Pi' = \mathcal{L}(\chi M)$ is supported on $C^{\beta'}(\mathcal{O})$ for any $\beta' < \alpha - d/2$, and its RKHS $\mathcal{H} = \{\chi F, F \in H^\alpha(\mathcal{O})\}$ is continuously imbedded into $H_c^\alpha(\mathcal{O})$ and contains $H_K^\alpha(\mathcal{O})$. The condition $F_0 \in \mathcal{H}$ then equals to the standard assumption that $F_0 \in H^\alpha(\mathcal{O})$ is supported on a strict subset K of \mathcal{O} .

In the following we consider the re-scaled prior

$$\Pi_N = \mathcal{L}(F_N), \quad F_N = N^{-\frac{d}{4\alpha+8+2d}} F', \quad (15)$$

where $F' \sim \Pi'$. Then Π_N defines a centred Gaussian prior on $C(\mathcal{O})$, and its RKHS \mathcal{H}_N is given by \mathcal{H} with the norm

$$\|F\|_{\mathcal{H}_N} = N^{\frac{d}{4\alpha+8+2d}} \|F\|_{\mathcal{H}}.$$

Theorem 3. For a fixed integer $\alpha > \beta + d/2$, $\beta \geq 2$, consider the Gaussian prior Π_N in (15) with the base prior Π' satisfying condition 2 with RKHS \mathcal{H} . Let $\Pi_N(\cdot | Y^N, Z^N)$ be the resulting posterior distribution arising from observations (Y^N, Z^N) in (10) with the boundary and initial value functions in (2) satisfying (12)–(14), and with $F = F_0 \in \mathcal{H}$. Set $\delta_N = N^{-(\alpha+2)/(2\alpha+4+d)}$.

Then for any $D > 0$ there exists a sufficiently large $L > 0$ such that, as $N \rightarrow \infty$,

$$\Pi_N(F : \|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2(\mathcal{O})} > L\delta_N | Y^N, Z^N) = O_{P_{F_0}^N} \left(e^{-DN\delta_N^2} \right) \quad (16)$$

and for sufficiently large $M > 0$

$$\Pi_N(F : \|F\|_{C^\beta(\mathcal{O})} > M | Y^N, Z^N) = O_{P_{F_0}^N} \left(e^{-DN\delta_N^2} \right). \quad (17)$$

Next we will formulate a theorem about the posterior contraction around f_0 in L^2 -norm. For this we need the following push-forward posterior distribution

$$\tilde{\Pi}_N(\cdot | Y^N, Z^N) = \mathcal{L}(f), \quad f = \Phi \circ F, \quad F \sim \Pi_N(\cdot | Y^N, Z^N). \quad (18)$$

Theorem 4. Let $\Pi_N(\cdot | Y^N, Z^N)$, δ_N and F_0 be as in theorem 3 with an integer $\beta \geq 2$, and denote $f_0 = \Phi \circ F_0$. Then for any $D > 0$ there exists $L > 0$ large enough such that, as $N \rightarrow \infty$,

$$\tilde{\Pi}_N \left(f : \|f - f_0\|_{L^2(\mathcal{O})} > L\delta_N^{\frac{\beta}{2+\beta}} | Y^N, Z^N \right) = O_{P_{f_0}^N} \left(e^{-DN\delta_N^2} \right).$$

We can also show that the posterior mean $\mathbb{E}^\Pi(F | Y^N, Z^N)$ of $\Pi_N(\cdot | Y^N, Z^N)$ converges to F_0 with speed $\delta_N^{\frac{\beta}{2+\beta}}$.

Theorem 5. *Under the assumptions of theorem 4 let $\bar{F}_N = \mathbb{E}^\Pi(F | Y^N, Z^N)$ be the mean of $\Pi_N(\cdot | Y^N, Z^N)$. Then, as $N \rightarrow \infty$, we have*

$$P_{F_0}^N \left(\|\bar{F}_N - F_0\|_{L^2(\mathcal{O})} > L\delta_N^{\frac{\beta}{2+\beta}} \right) \rightarrow 0.$$

Note that, since a composition with the link function Φ is L^2 -Lipschitz, the above result also holds for the original potential f , that is, we can replace $\|\bar{F}_N - F_0\|_{L^2}$ by $\|\Phi \circ \bar{F}_N - f_0\|_{L^2}$. The proof of theorem 5 follows directly from theorems 3 and 4, and the proof of [27, theorem 6].

3.2. Truncated Gaussian priors

In practice so called sieve-priors, which are concentrated on a finite-dimensional approximation of the parameter space supporting the prior, are often employed for computational reasons. One of the commonly used methods is to use the truncated Karhunen–Loève series expansion of the Gaussian base prior Π' . The contraction rate (16) of the forward problem remains valid with these truncated priors if the approximation spaces are appropriately chosen. In this section we will establish the optimal rate of estimating f_0 from the data and show that it can be achieved with truncated and rescaled priors.

Let $\{\Psi_{l,r}, l \geq -1, r \in \mathbb{Z}^d\}$ be an orthonormal basis of $L^2(\mathbb{R}^d)$ composed of sufficiently regular, compactly supported Daubechies wavelets, see the proof of theorem 6 for more details. We assume that $F_0 \in H_K^\alpha(\mathcal{O})$ for some $K \in \mathcal{O}$, and denote by \mathcal{R}_l the set of indices r such that the support of $\Psi_{l,r}$ intersects with K . Fix any compact $K' \subset \mathcal{O}$ such that $K \subsetneq K'$ and a cut-off function $\chi \in C_c^\infty(\mathcal{O})$ for which $\chi = 1$ on K' . Let $\alpha > 2 + d/2$, and consider the prior Π'_J arising as the law of the Gaussian sum

$$\Pi'_J = \mathcal{L}(\chi\tilde{F}), \quad \tilde{F} = \sum_{\substack{l \leq J \\ r \in \mathcal{R}_l}} 2^{-\alpha l} F_{l,r} \Psi_{l,r}, \quad F_{l,r} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1), \quad (19)$$

where $J = J_N \rightarrow \infty$ is a deterministic truncation point. Then Π'_J defines a centred Gaussian prior that is supported on a finite dimensional space

$$\mathcal{H}_J := \text{span}\{\chi\Psi_{l,r}, l \leq J, r \in \mathcal{R}_l\} \subset C(\mathcal{O}).$$

Theorem 6. *Let Π_N be the rescaled prior as in (15), where now $F' \sim \Pi'_J$, with Π'_J defined in (19), and $J = J_N \in \mathbb{N}$ is chosen so that $2^J \simeq N^{1/(2\alpha+4+d)}$. Let $\Pi_N(\cdot | Y^N, Z^N)$ be the resulting posterior distribution arising from data (Y^N, Z^N) in (10) with the boundary and initial value functions in (2) satisfying (12)–(14), with $F = F_0 \in H_K^\alpha(\mathcal{O})$. Let δ_N be as in theorem 3, and assume that $f_0 = \Phi \circ F_0$. Then (16) remains valid and for any $D > 0$, and for sufficiently large $M > 0$,*

$$\Pi_N \left(F : \|F\|_{H^\alpha(\mathcal{O})} > M | Y^N, Z^N \right) = O_{P_{F_0}^N} \left(e^{-DN\delta_N^2} \right), \quad (20)$$

as $N \rightarrow \infty$. Furthermore, let $\tilde{\Pi}_N(\cdot | Y^N, Z^N)$ be the push-forward posterior as in (18). Then there exists $L > 0$ large enough such that

$$\tilde{\Pi}_N \left(f : \|f - f_0\|_{L^2(\mathcal{O})} > LN^{-\frac{\alpha}{2\alpha+4+d}} | Y^N, Z^N \right) = O_{P_{f_0}^N} \left(e^{-DN\delta_N^2} \right), \quad (21)$$

as $N \rightarrow \infty$.

Theorem 7. Under the assumptions of theorem 6 let $\bar{F}_N = \mathbb{E}^\Pi(F | Y^N, Z^N)$ be the mean of $\Pi_N(\cdot | Y^N, Z^N)$. Then, as $N \rightarrow \infty$, we have

$$P_{f_0}^N \left(\|\Phi \circ \bar{F}_N - f_0\|_{L^2(\mathcal{O})} > LN^{-\frac{\alpha}{2\alpha+4+d}} \right) \rightarrow 0.$$

The proof of theorem 7 follows directly from theorem 6 and the proof of [27, theorem 6]. Note that from a non-asymptotic point of view, the N -dependent rescaling of the prior can be thought just as an adjustment of the covariance operator of the prior. Similar result to theorems 6 and 7 could be proven for more general Gaussian priors which are not of wavelet type but we do not pursue these extensions in this paper.

If we compare the rates achieved in section 3.1 to those attained in this section we notice that while the rates for Gaussian process priors are of order $\delta_N^{\beta/(2+\beta)}$ the rates achieved for truncated Gaussian priors are $\delta_N^{\alpha/(2+\alpha)}$ with $\alpha > \beta + d/2$. The question then arises as to what is the optimal rate for estimating f from the data. Next we will give a minimax lower bound on the rate of estimation for f and note that the rates attained in this section are minimax optimal. We also note that, by modifying the proof of theorem 8, one can show that the rate δ_N achieved in (16) for the PDE-constrained regression problem of recovering $\mathcal{G}(F_0)$ in prediction loss is minimax optimal. Notice that the following theorem gives a lower bound that holds for any estimate for f , not just the ones studied in this paper.

Theorem 8. Let f be the absorption term in (2) and let g, u_0 satisfy (12)–(14) with α replaced by $\alpha + d/2$, with $\alpha > 2 + d/2$. Then there exists $c > 0$ such that for $\epsilon > 0$ arbitrarily small, as $N \rightarrow \infty$,

$$\liminf_{\hat{f}_N} \sup_{f \in \tilde{\mathcal{F}}_\alpha} P_f^N \left(\|\hat{f}_N - f\|_{L^2(\mathcal{O})} > cN^{-\frac{\alpha}{2\alpha+4+d}} \right) \geq 1 - \epsilon, \quad (22)$$

where $\tilde{\mathcal{F}}_\alpha = \{f \in C^\alpha(\mathcal{O}) : \inf_{x \in \mathcal{O}} f(x) \geq f_{\min} > 0, \|f\|_{C^\alpha} \leq B\}$, with any sufficient large $B > 0$, and the infimum is taken over all measurable functions $\hat{f}_N = \hat{f}(Y^N, Z^N, g, u_0)$, where the observations (Y^N, Z^N) are generated through model (8).

To prove theorem 8 we use [25, theorem 6.3.2] which allows us to reduce the problem of estimating the lower bound (22) into calculating lower bounds of certain testing problems that involve several hypotheses in the parameter space. We then use tools from information theory, and in particular the Kullback–Leibler distance between two probability measures, to attain lower bounds for these testing problems.

As mentioned in the introduction the posterior distributions arising from Gaussian priors can be computed using MCMC algorithms [12, 17]. These algorithms often employ a finite-dimensional approximation of the parameter space as in this section. Non-asymptotic sampling bounds for the preconditioned Crank–Nicholson algorithm were established in [28] providing bounds on the approximation error for the computation of the posterior mean. Note that these bounds hold even for likelihood functions that are not log-concave as in this paper. See also [14, 53] for recent result on efficiently generating random samples from high-dimensional and non-log-concave posterior measures.

3.3. Conclusions

The convergence rates obtained in this article demonstrate the frequentist consistency of the Bayesian inversion with Gaussian process priors in the parabolic inverse problem (2) with data (10) in the large sample limit $N \rightarrow \infty$. The rates in section 3.1 for the rescaled Gaussian process prior are minimax optimal for the PDE-constrained regression problem of recovering $\mathcal{G}(F_0)$ in prediction loss. However, even though the rates for estimating f_0 approach the optimal rate $N^{-1/2}$ for very smooth models they are not optimal in general. We note that the optimal rate in theorem 8 equals to $\delta_N^{\alpha/(2+\alpha)}$ and so the rates achieved in theorems 4 and 5 would be optimal if we could replace β by $\alpha > \beta + d/2$. The second part of theorem 3 with the proof of theorem 4 reveals that this suboptimal rate is due to the fact that, while we are interested in recovering $F_0 \in \mathcal{H} \subset H_c^\alpha$, we can only show that the posterior mass is concentrated in C^β -balls, with $\beta < \alpha - d/2$. On the other hand, we can show that the posterior mass of the truncated Gaussian priors is concentrated in H^α -balls which allows us to attain the optimal convergence rate. We note that truncation of a Gaussian prior increases its regularity, which can correct for possible under-smoothing. The question of optimality for infinite dimensional Gaussian priors remains an interesting avenue for future research in non-linear inverse problems.

A natural question that arises in section 3.2 is whether the non-linear inverse problem considered can be solved in a fully Bayesian way (prior independent of the measurement). One way of achieving this would be to employ a hierarchical prior that randomises the finite truncation point J in the Karhunen–Loève series expansion (19). Such an approach has been investigated in [27] for an elliptic PDE with similar consistency results as in theorems 3–5 for smooth enough ground truth. Another avenue for further research is assuming the noise amplitude σ to be unknown and using a hierarchical prior for it.

4. Proofs

4.1. Proofs of the main results

The proof of theorem 5 follows directly from theorems 3 and 4, and the proof of [27, theorem 6]. The proofs of theorems 3 and 4 rely on the following forward and stability estimates. The proofs of the propositions can be found in section 4.2.

Proposition 9. *Let \mathcal{G} be the solution map defined in (9) with g, u_0 as in (12) and (13). Let $\alpha > 2 + d/2$ and $F_1, F_2 \in H_c^\alpha(\mathcal{O})$. Then*

$$\|\mathcal{G}(F_1) - \mathcal{G}(F_2)\|_{L^2(\mathcal{Q})} \lesssim (1 + \|F_1\|_{C^2(\mathcal{O})}^4 \vee \|F_2\|_{C^2(\mathcal{O})}^4) \|F_1 - F_2\|_{(H^2(\mathcal{O}))^*},$$

where X^* denotes the topological dual space of a normed linear space X . Also, there exists $C > 0$ such that

$$\sup_{F \in H_c^\alpha(\mathcal{O})} \|\mathcal{G}(F)\|_\infty \leq C(\|g\|_\infty + \|u_0\|_\infty) < \infty.$$

Proof of theorem 3. It follows from proposition 9 that the inverse problem (10) falls in the general framework studied in [27], with $\beta = 2$, $\gamma = 4$, $\kappa = 2$ and $S = c(\|g\|_\infty + \|u_0\|_\infty)$. Theorem 3 then follows directly from [27, theorem 14]. \square

Proposition 10. Let $G(f)$ be the solution to (2) with g, u_0 as in (12) and (13), and $f, f_0 \in \mathcal{F}_{\alpha, f_{\min}}$. Then

$$\|f - f_0\|_{L^2(\mathcal{O})} \lesssim e^{c\|f \vee f_0\|_{\infty}} \|G(f) - G(f_0)\|_{H^{2,1}(\mathcal{Q})}.$$

To prove theorem 4 we will also need that the forward map G maps bounded sets in \mathcal{C}^{β} onto bounded sets in $H^{2+\beta, 1+\beta/2}(\mathcal{Q})$.

Proposition 11. Let $\beta > 0$ and $f \in \mathcal{C}^{\beta}(\mathcal{O})$, with $\inf_{x \in \mathcal{O}} f(x) \geq f_{\min} > 0$. Then there exists a constant $C > 0$ such that

$$\|G(f)\|_{H^{2+\beta, 1+\beta/2}(\mathcal{Q})} \leq C(1 + \|f\|_{\mathcal{C}^{\beta}}^{1+\beta/2}).$$

Proof of theorem 4. We start by noting that if $\Phi: \mathbb{R} \rightarrow \mathbb{R}$ is a regular link function in the sense of definition 1 then for each integer $m \geq 0$ there exists $C > 0$ such that, for all $F \in C^m(\mathcal{O})$,

$$\|\Phi \circ F\|_{C^m} \leq C(1 + \|F\|_{C^m}^m). \quad (23)$$

See [52, lemma 29] for proof. With the above bound we can use the conclusion of theorem 3 for the push-forward posterior $\tilde{\Pi}_N(\cdot | Y^N, Z^N)$. Estimate (16) directly implies that

$$\tilde{\Pi}_N(f: \|G(f) - G(f_0)\|_{L^2} > L\delta_N | Y^N, Z^N) = O_{P_{f_0}^N}(e^{-DN\delta_N^2}),$$

as $N \rightarrow \infty$. Using the bound (23) and the estimate (17) we get, for a sufficiently large $M' > 0$,

$$\tilde{\Pi}_N(f: \|f\|_{\mathcal{C}^{\beta}} > M' | Y^N, Z^N) \leq \Pi_N(F: \|F\|_{\mathcal{C}^{\beta}} > M | Y^N, Z^N) = O_{P_{f_0}^N}(e^{-DN\delta_N^2}).$$

Since $\mathcal{C}^{\beta} \subset \mathcal{C}^{\beta}$ the above estimate is still true if we replace $\|\cdot\|_{\mathcal{C}^{\beta}}$ by $\|\cdot\|_{\mathcal{C}^{\beta}}$.

If $f \in \mathcal{C}^{\beta}$ with $\|f\|_{\mathcal{C}^{\beta}} \leq M'$ proposition 11 implies that $G(f), G(f_0) \in H^{2+\beta, 1+\beta/2}$ and

$$\|G(f_0)\|_{H^{2+\beta, 1+\beta/2}} \lesssim 1 + \|f_0\|_{\mathcal{C}^{\beta}}^{1+\beta/2} < \infty, \|G(f)\|_{H^{2+\beta, 1+\beta/2}} \lesssim 1 + \|f\|_{\mathcal{C}^{\beta}}^{1+\beta/2} \leq M'' < \infty.$$

We will also need the following interpolation inequality. Let $s \geq 0$ and $\theta \in (0, 1)$. Then for all $u \in H^{s, s/2}(\mathcal{Q})$

$$\|u\|_{H^{(1-\theta)s, (1-\theta)s/2}} \lesssim \|u\|_{H^{s, s/2}}^{1-\theta} \|u\|_{L^2}^{\theta}, \quad (24)$$

see [42, chapter 4, proposition 2.1]. Combining the above we see that

$$\begin{aligned} \|G(f) - G(f_0)\|_{H^{2,1}} &\lesssim \|G(f) - G(f_0)\|_{H^{2+\beta, 1+\beta/2}}^{\frac{2}{2+\beta}} \|G(f) - G(f_0)\|_{L^2}^{\frac{\beta}{2+\beta}} \\ &\lesssim \|G(f) - G(f_0)\|_{L^2}^{\frac{\beta}{2+\beta}}. \end{aligned}$$

Hence we get, for large enough $L > 0$,

$$\begin{aligned} & \tilde{\Pi}_N \left(f : \|G(f) - G(f_0)\|_{H^{2,1}} > L\delta_N^{\frac{\beta}{2+\beta}} \mid Y^N, Z^N \right) \\ & \leq \tilde{\Pi}_N(f : \|G(f) - G(f_0)\|_{L^2} > L\delta_N \mid Y^N, Z^N) + \tilde{\Pi}_N(f : \|f\|_{C^\beta} > M' \mid Y^N, Z^N) \\ & = O_{P_{f_0}^N}(e^{-ND\delta_N^2}), \end{aligned}$$

as $N \rightarrow \infty$. Applying the stability estimate of proposition 10 we can then conclude

$$\begin{aligned} & \tilde{\Pi}_N \left(f : \|f - f_0\|_{L^2} > L\delta_N^{\frac{\beta}{2+\beta}} \mid Y^N, Z^N \right) \\ & \leq \tilde{\Pi}_N \left(f : \|G(f) - G(f_0)\|_{H^{2,1}} > L'\delta_N^{\frac{\beta}{2+\beta}} \mid Y^N, Z^N \right) + \tilde{\Pi}_N(f : \|f\|_{C^\beta} > M' \mid Y^N, Z^N) \\ & = O_{P_{f_0}^N}(e^{-ND\delta_N^2}). \end{aligned}$$

□

Proof of theorem 6. We start by introducing the wavelet setting that is used in the proofs of theorems 6 and 8. Consider an s -regular orthonormal wavelet basis for the Hilbert space $L^2(\mathbb{R}^d)$ given by compactly supported Daubechies tensor wavelet basis functions

$$\{\Psi_{l,k} : l \in \mathbb{N} \cup \{-1, 0\}, k \in \mathbb{Z}^d\} \quad \Psi_{l,k} = 2^{\frac{ld}{2}} \Psi_{0,k}(2^l \cdot), \quad \text{for } l \geq 0.$$

Note that we can choose the smoothness s of the basis to be as large as required and hence will omit it in what follows. For more details about wavelets see [19, 45] or [25, chapter 4]. For $\alpha \in \mathbb{R}$ we have the following wavelet characterisation of $H^\alpha(\mathbb{R}^d)$ norm

$$\|f\|_{H^\alpha(\mathbb{R}^d)}^2 \simeq \sum_{l,k} 2^{2l\alpha} \langle f, \Psi_{l,k} \rangle_{L^2(\mathbb{R}^d)}^2. \tag{25}$$

We also note that, for $\alpha \geq 0$ and some $C > 0$,

$$f \in C^\alpha(\mathbb{R}^d) \Rightarrow \sup_{l,k} 2^{l(\alpha+d/2)} |\langle f, \Psi_{l,k} \rangle_{L^2(\mathbb{R}^d)}| \leq C \|f\|_{C^\alpha(\mathbb{R}^d)},$$

with the converse being true when $\alpha \notin \mathbb{N}$.

We can construct an orthonormal wavelet basis of $L^2(\mathcal{O})$ given by

$$\{\Psi_{l,k}^\mathcal{O} : k \leq N_l, l \in \mathbb{N} \cup \{-1, 0\}\}, \quad N_l \in \mathbb{N},$$

that consists of interior wavelets $\Psi_{l,k}^\mathcal{O} = \Psi_{l,k}$, which are compactly supported in \mathcal{O} , and of boundary wavelets $\Psi_{l,k}^\mathcal{O} = \Psi_{l,k}^b$, which are an orthonormalised linear combination of those wavelets that have support inside and outside \mathcal{O} , see [60, theorem 2.33]. Using the above basis any function $f \in L^2(\mathcal{O})$ has orthogonal wavelet series expansion

$$f = \sum_l \sum_{k=1}^{N_l} \langle f, \Psi_{l,k}^\mathcal{O} \rangle_{L^2(\mathcal{O})} \Psi_{l,k}^\mathcal{O}.$$

We denote by $F^J = P_{\mathcal{H}_J}(F) \in \mathcal{H}_J$ the wavelet projection (26) and note that for all $F_0 \in H_K^\alpha(\mathcal{O})$

$$\|F_0 - F_0^J\|_{(H^2(\mathcal{O}))^*} \lesssim 2^{-J(\alpha+2)}$$

by (63) from [27]. We also note that $\|F_0^J\|_{C^2} \leq \|F_0\|_{C^2}$ by standard properties of wavelet bases and hence, choosing $2^J = N^{1/(2\alpha+4+d)}$, it follows from proposition 9 that

$$\|\mathcal{G}(F_0) - \mathcal{G}(F_0^J)\|_{L^2(\mathcal{Q})} \lesssim \|F_0 - F_0^J\|_{(H^2(\mathcal{O}))^*} \lesssim N^{-\frac{\alpha+2}{2\alpha+4+d}} = \delta_N.$$

Using triangle inequality we then see that for a sufficiently large $c > 0$

$$\Pi_N(\|\mathcal{G}(F) - \mathcal{G}(F_0)\|_{L^2(\mathcal{Q})} > c\delta_N) \geq \Pi_N(\|\mathcal{G}(F) - \mathcal{G}(F_0^J)\|_{L^2(\mathcal{Q})} > c'\delta_N).$$

We can then conclude that the results of theorems 3–5 remain valid under the truncated and rescaled Gaussian prior, see [27] section 3.2.

Following the idea from the proof of theorem 4.13 [53] we will next show that the posterior mass concentrates in some H^α -balls. Define for any $M', Q > 0$

$$\mathcal{A}_N = \{F = F_1 + F_2 \in \mathcal{H}_J : \|F_1\|_{(H^2)^*} \leq Q\delta_N, \|F_2\|_{\mathcal{H}} \leq M'\}.$$

Then, by theorem 13 with lemmas 17 and 18 in [27], we have for Q, M' sufficiently large

$$\Pi_N(F \in \mathcal{A}_N | Y^N, Z^N) \geq 1 - O_{P_{F_0}^N}(e^{-DN\delta_N^2}).$$

To prove (20) we need to show that $\|F\|_{H^\alpha} \leq M$ for all $F \in \mathcal{A}_N$.

Let $\alpha \geq 0$. Then

$$\tilde{F} = \sum_{\substack{l \leq J \\ r \in \mathcal{R}_l}} F_{l,r} \Psi_{l,r} = \sum_{\substack{l \leq J \\ r \in \mathcal{R}_l}} 2^{-\alpha l} G_{l,r} \Psi_{l,r}, \quad G_{l,r} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1),$$

defines a centred Gaussian probability measure supported on $\tilde{\mathcal{H}}_J = \text{span}\{\Psi_{l,r}, l \leq J, r \in \mathcal{R}_l\}$ with the RKHS $\tilde{\mathcal{H}}_J$ endowed with norm

$$\|\tilde{F}\|_{\tilde{\mathcal{H}}_J}^2 = \sum_{\substack{l \leq J \\ r \in \mathcal{R}_l}} 2^{2\alpha} F_{l,r}^2 = \|\tilde{F}\|_{H^\alpha(\mathbb{R}^d)}^2 \quad \forall \tilde{F} \in \tilde{\mathcal{H}}_J.$$

The random function

$$F = \chi \tilde{F} = \sum_{\substack{l \leq J \\ r \in \mathcal{R}_l}} F_{l,r} \chi \Psi_{l,r} = \sum_{\substack{l \leq J \\ r \in \mathcal{R}_l}} 2^{-\alpha l} G_{l,r} \chi \Psi_{l,r}, \quad G_{l,r} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1),$$

then defines the centred Gaussian probability measure Π'_J , as in (19), supported on $\mathcal{H}_J = \text{span}\{\chi \Psi_{l,r}, l \leq J, r \in \mathcal{R}_l\}$, with the RKHS norm satisfying

$$\|F\|_{\mathcal{H}_J}^2 = \|\chi \tilde{F}\|_{\mathcal{H}_J}^2 \leq \|\tilde{F}\|_{\tilde{\mathcal{H}}_J}^2 = \sum_{\substack{l \leq J \\ r \in \mathcal{R}_l}} 2^{2\alpha} F_{l,r}^2 \quad \forall F \in \mathcal{H}_J.$$

We also note that for all $\tilde{F} \in \tilde{\mathcal{H}}_J$ there exists $\tilde{F}' \in \tilde{\mathcal{H}}_J$ such that $\chi\tilde{F} = \chi\tilde{F}'$ and

$$\|\chi\tilde{F}\|_{\mathcal{H}_J} = \|\tilde{F}'\|_{\tilde{\mathcal{H}}_J}.$$

Hence, if $F = \chi\tilde{F}$ is an arbitrary element of \mathcal{H}_J we can write

$$\|F\|_{H^\alpha(\mathcal{O})} = \|\chi\tilde{F}'\|_{H^\alpha(\mathcal{O})} \leq \|\tilde{F}'\|_{H^\alpha(\mathbb{R}^d)} = \|\tilde{F}'\|_{\tilde{\mathcal{H}}_J} = \|F\|_{\mathcal{H}_J} \quad \forall F \in \mathcal{H}_J.$$

We will next show that

$$\|F\|_{H^\alpha(\mathcal{O})} \lesssim 2^{J(\alpha+2)} \|F\|_{(H^2(\mathcal{O}))^*} \quad \forall F \in \mathcal{H}_J.$$

Denote for $J' \in \mathbb{N}$, $J' \leq J$,

$$F^{J'} = P_{\mathcal{H}_{J'}}(F) = \sum_{\substack{l \leq J' \\ r \in \mathcal{R}_l}} \langle F, \Psi_{l,r} \rangle \chi \Psi_{l,r} \in \mathcal{H}_{J'}. \quad (26)$$

Note that for large enough $J_{\min} \in \mathbb{N}$, if $l \geq J_{\min}$ and the support of $\Psi_{l,r}$ intersects K , then $\text{supp}(\Psi_{l,r}) \subset K'$ and we have $\chi\Psi_{l,r} = \Psi_{l,r}$ for all $l \geq J_{\min}$ and $r \in \mathcal{R}_l$. We can then write

$$F = F^{J_{\min}} + (F - F^{J_{\min}}) = \sum_{\substack{l \leq J_{\min} \\ r \in \mathcal{R}_l}} F_{l,r} \chi \Psi_{l,r} + \sum_{\substack{J_{\min} < l \leq J \\ r \in \mathcal{R}_l}} F_{l,r} \Psi_{l,r}.$$

Since $\mathcal{H}_{J_{\min}}$ is a fixed finite dimensional subspace we get, by equivalence of norms, that $\|F^{J_{\min}}\|_{H^\alpha} \leq C_{J_{\min}} \|F^{J_{\min}}\|_{(H^2)^*} \leq C_{J_{\min}} \|F\|_{(H^2)^*}$. We also see that $F - F^{J_{\min}}$ is compactly supported on \mathcal{O} and hence can be extended by zero to \mathbb{R}^d . We can then write

$$\begin{aligned} \|F - F^{J_{\min}}\|_{H^\alpha(\mathcal{O})}^2 &= \sum_{\substack{J_{\min} < l \leq J \\ r \in \mathcal{R}_l}} 2^{2l\alpha} F_{l,r}^2 \\ &= \sum_{\substack{J_{\min} < l \leq J \\ r \in \mathcal{R}_l}} 2^{2l\alpha+4l} 2^{-4l} F_{l,r}^2 \\ &\leq 2^{J(2\alpha+4)} \|F - F^{J_{\min}}\|_{(H^2(\mathcal{O}))^*}^2 \\ &\leq 2^{J(2\alpha+4)} \|F\|_{(H^2(\mathcal{O}))^*}^2. \end{aligned}$$

Combining the above we see that, for $F \in \mathcal{H}_J$, $\|F\|_{(H^2)^*} \leq Q\delta_N$ and $2^J = N^{1/(2\alpha+4+d)}$,

$$\|F\|_{H^\alpha(\mathcal{O})} \leq 2^{J(\alpha+2)} \|F\|_{(H^2(\mathcal{O}))^*} \leq QN^{\frac{\alpha+2}{2\alpha+4+d}} \delta_N = Q,$$

which concludes the first part of the proof.

To prove the optimal convergence rate we need to replace the $\|\cdot\|_{C^\beta}$ -bounds in the proof of theorem 4 by $\|\cdot\|_{H^\alpha}$ -bounds. To do this we note that if $\alpha > d/2$ and $f \in H^\alpha(\mathcal{O})$

with $\inf_{x \in \mathcal{O}} f(x) \geq f_{\min} > 0$ we can use the inequality (3) instead of (4) in the proof of proposition 11 and show that

$$\|G(f)\|_{H^{2+\alpha, 1+\alpha/2}(Q)} \leq C(1 + \|f\|_{H^\alpha}^{1+\alpha/2}).$$

Also, if $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ is a regular link function in the sense of definition 1 then for each integer $m \geq 0$ there exists $C > 0$ such that, for all $F \in H^m(\mathcal{O})$,

$$\|\Phi \circ F\|_{H^m} \leq C(1 + \|F\|_{H^m}^m). \quad (27)$$

See [52, lemma 29] for proof. The result then follows directly from the proof of theorem 4. \square

Proof of theorem 8. The proof uses similar ideas to [49, 52] by applying [25, theorem 6.3.2] which reduces the problem of estimating the lower bound in the whole parameter space into a testing problem in a finite subset $(f_m : m = 0, \dots, M)$ of $\tilde{\mathcal{F}}_\alpha$. First, note that

$$\liminf_{\hat{f}_N} \sup_{f \in \tilde{\mathcal{F}}_\alpha} P_f^N \left(\|\hat{f}_N - f\|_{L^2} > r_N \right) \geq \liminf_{\hat{f}_N} \max_{m=0, \dots, M} P_{f_m}^N \left(\|\hat{f}_N - f_m\|_{L^2} > r_N \right) \quad (28)$$

for any finite set $(f_m : m = 0, \dots, M)$ in $\tilde{\mathcal{F}}_\alpha$. We can use any estimator \hat{f}_N to test between the $M + 1$ hypothesis by choosing the f_m closest to \hat{f}_N , that is, we are looking for Ψ_N such that

$$\|\hat{f}_N - f_{\Psi_N}\|_{L^2} = \min_{m=0, \dots, M} \|\hat{f}_N - f_m\|_{L^2}.$$

The errors of the test, which are given by the probability that even though the measurement was generated by f_m the estimator \hat{f}_N is closest to $f_{m'}$ with some $m' \neq m$, are bounded by

$$P_{f_m}^N (\Psi_N \neq m) \leq P_{f_m}^N \left(\|\hat{f}_N - f_{m'}\|_{L^2} \leq \|\hat{f}_N - f_m\|_{L^2} \quad \text{for some } m' \right).$$

Using triangle inequality we see that if $\|\hat{f}_N - f_{m'}\|_{L^2} \leq \|\hat{f}_N - f_m\|_{L^2}$ for some m' ,

$$\|\hat{f}_N - f_m\|_{L^2} \geq \|f_m - f_{m'}\|_{L^2} - \|\hat{f}_N - f_{m'}\|_{L^2} \geq \|f_m - f_{m'}\|_{L^2} - \|\hat{f}_N - f_m\|_{L^2}. \quad (29)$$

Assume next that the f_m are $2r_N$ separated, that is, $\|f_m - f_{m'}\|_{L^2} \geq 2r_N$ for all $m \neq m'$. Then (29) implies

$$\|\hat{f}_N - f_m\|_{L^2} \geq r_N$$

and we see that

$$P_{f_m}^N (\Psi_N \neq m) \leq P_{f_m}^N \left(\|\hat{f}_N - f_m\|_{L^2} \geq r_N \right).$$

We can then conclude using (28) that

$$\liminf_{\hat{f}_N} \sup_{f \in \tilde{\mathcal{F}}_\alpha} P_f^N \left(\|\hat{f}_N - f\|_{L^2} > r_N \right) \geq \liminf_{\Psi_N} \max_{m=0, \dots, M} P_{f_m}^N (\Psi_N \neq m)$$

if the hypothesis $(f_m : m = 0, \dots, M)$ are $2r_N$ separated.

If we can find $(f_m : m = 0, \dots, M_N) \in \tilde{\mathcal{F}}_\alpha$ that are $r_N \simeq N^{-\alpha/(2\alpha+4+d)}$ separated from each other and show that, for some $\epsilon > 0$ and M_N ,

$$\text{KL}(P_{f_m}^N, P_{f_0}^N) \leq \epsilon \log(M_N),$$

where KL denotes the Kullback–Leibler divergence, we can use [25, theorem 6.3.2] which states that

$$\liminf_{\Psi_N} \max_{m=0, \dots, M} P_{f_m}^N(\Psi_N \neq m) \geq \frac{\sqrt{M_N}}{1 + \sqrt{M_N}} \left(1 - 2\epsilon - \sqrt{\frac{8\epsilon}{\log M_N}} \right).$$

(a) We will start by showing that \tilde{F}_α contains $\{f_m : m = 0, 1, \dots, M\}, M \geq 1$ such that

$$\|f_m - f_{m'}\|_{L^2} \gtrsim N^{-\frac{\alpha}{2\alpha+4+d}} \quad \text{for all } m \neq m',$$

that is, f_m are $N^{-\frac{\alpha}{2\alpha+4+d}}$ -separated from each other.

We use the wavelet basis setting described at the beginning of the proof of theorem 6. For every $j \in \mathbb{N}$ there exists a small positive constant c and $n_j = c2^{jd}$ many Daubechies wavelets $(\Psi_{jr} : r = 1, \dots, n_j)$ with disjoint compact supports in \mathcal{O} . Let $b_{m,\cdot}$ be a point in the discrete hypercube $\{-1, 1\}^{n_j}$. We define

$$h_m(x) = h_{m,j}(x) = \kappa \sum_{r=1}^{n_j} b_{m,r} 2^{-j(\alpha+d/2)} \Psi_{jr}(x), \quad x \in \mathcal{O}, \tag{30}$$

where κ can be chosen to be as small as wanted, and $m = 0, \dots, M_j$ with M_j chosen later. Note that h_m is compactly supported in \mathcal{O} and has zero extension from \mathcal{O} to \mathbb{R}^d with the global Hölder norm being equal to the intrinsic one. For $\alpha \in \mathbb{N}$ we can write

$$\begin{aligned} \sup_{x \in \mathbb{R}^d} |D^\alpha h_m(x)| &= \sup_{x \in \mathcal{O}} \left| \kappa \sum_{r=1}^{n_j} b_{m,r} 2^{-j(\alpha+d/2)} D^\alpha \Psi_{jr}(x) \right| \\ &\leq \kappa \sup_{x \in \mathcal{O}} \sum_{r=1}^{n_j} |(D^\alpha \Psi_{0,r})(2^j x)| \leq C\kappa, \end{aligned}$$

where the last inequality follows from the fact that at any point there are only finitely many $\Psi_{0,r}$ that get a non-zero value. Since the interior wavelets are orthogonal to the boundary wavelets we get for $\alpha \notin \mathbb{N}$

$$\|h_m\|_{C^\alpha} \leq C \sup_{l,k} 2^{l(\alpha+d/2)} |\langle h_m, \Psi_{l,k} \rangle_{L^2}| = C\kappa.$$

Hence, by choosing κ small enough, we see that all h_m are contained in $\{h \in C^\alpha(\mathcal{O}) : \|h\|_{C^\alpha} \leq 1\}$.

Let $f_0 \equiv 1, h_m$ as in (30), and define functions

$$f_m = f_0 + h_m, \quad m = 1, \dots, M_j.$$

We then have $\|f_m\|_{C^\alpha} \leq \|f_0\|_{C^\alpha} + C\kappa$, and for κ small enough f_m is bounded away from zero. By the Varshamov–Gilbert bound there exists $\{b_{m,\cdot} : m = 1, \dots, M_j\} \in \{-1, 1\}^{n_j}$,

with $M_j \geq 3^{\frac{n_j}{4}}$, that are $n_j/8$ -separated in the Hamming-distance, that is, for all $m, m' \leq M_j$ and $m \neq m'$

$$\sum_{r=1}^{n_j} (b_{m,r} - b_{m',r})^2 \gtrsim n_j$$

see e.g. [25, example 3.1.4]. Setting $2^j \simeq N^{\frac{1}{2\alpha+4+d}}$ we get, for the f_m corresponding to the above separated $b_{m,\cdot}$, that

$$\begin{aligned} \|f_m - f_{m'}\|_{L^2}^2 &= \|h_m - h_{m'}\|_{L^2}^2 \\ &= \kappa^2 2^{-2j(\alpha+d/2)} \left\| \sum_{r=1}^{n_j} (b_{m,r} - b_{m',r}) \Psi_{j,r} \right\|_{L^2}^2 \\ &= \kappa^2 2^{-2j(\alpha+d/2)} \sum_{r=1}^{n_j} (b_{m,r} - b_{m',r})^2 \\ &\gtrsim \kappa^2 2^{-2j(\alpha+d/2)} n_j \gtrsim N^{\frac{2\alpha}{2\alpha+4+d}}. \end{aligned}$$

(b) Next we will prove that, for some $\epsilon > 0$,

$$\text{KL}(P_{f_m}^N, P_{f_0}^N) \leq \epsilon \log(M_j),$$

where KL denotes the Kullback–Leibler divergence.

Using (11) we see that

$$\begin{aligned} \mathbb{E}_{f_m}^i \left(\log \frac{dP_{f_m}^i(Y_i, Z_i)}{dP_{f_0}^i(Y_i, Z_i)} \right) &= \mathbb{E}_{f_m}^i \left(\frac{1}{2\sigma^2} ((Y_i - u_{f_0}(Z_i))^2 - (Y_i - u_{f_m}(Z_i))^2) \right) \\ &= \frac{1}{2\sigma^2} \mathbb{E}^\mu (u_{f_0}(Z_i)^2 - 2u_{f_0}(Z_i)u_{f_m}(Z_i) + u_{f_m}(Z_i)^2) \\ &\simeq \|u_{f_0} - u_{f_m}\|_{L^2}^2. \end{aligned}$$

Since $P_{f_m}^N$ is the product measure $\otimes_{i=1}^N P_F^i$ we have $\text{KL}(P_{f_m}^N, P_{f_0}^N) \simeq N \|u_{f_0} - u_{f_m}\|_{L^2}^2$. Using lemma 15 and (25) we then get

$$\begin{aligned} \|u_{f_m} - u_{f_0}\|_{L^2(Q)}^2 &\lesssim \|f_m - f_0\|_{(H_0^2(\mathcal{O}))^*}^2 \\ &\lesssim \|h_m\|_{H^{-2}(\mathbb{R}^d)}^2 \\ &= \kappa^2 2^{-2j(\alpha+d/2+2)} \sum_{r=1}^{n_j} 1 \lesssim \kappa^2 N^{-1} n_j. \end{aligned}$$

By the definition of M_j and choosing κ small enough we can conclude that $\text{KL}(P_{f_m}^N, P_{f_0}^N) \leq \epsilon \log(M_j)$.

Theorem 6.3.2 from [25] then states that

$$N^{\frac{\alpha}{2\alpha+4+d}} \inf_{\hat{f}_N} \sup_{f \in \tilde{\mathcal{F}}_\alpha} \mathbb{E}_f^N \|\hat{f}_N - f\|_{L^2(\mathcal{O})} \geq \frac{\sqrt{M_N}}{1 + \sqrt{M_N}} \left(1 - 2\epsilon - \sqrt{\frac{8\epsilon}{\log M_N}} \right),$$

where $M_N = M_j \rightarrow \infty$ when $N \rightarrow \infty$, and ϵ can be chosen to be as small as required by choosing κ small enough. We conclude the proof by noting that the proof of [25, theorem 6.3.2] actually states a slightly stronger result, showing the above lower bound for the testing problem $\inf_{\Psi_N} \max_{m=0, \dots, M} P_{f_m}^N (\Psi_N \neq m)$. \square

4.2. Proofs of the propositions

We start by proving some useful properties of the solutions to the inhomogeneous problem (31) below. Let $f \in C(\bar{\mathcal{O}})$ and $f > 0$. We denote by S_f the forward operator

$$S_f : H_{B,0}^{2,1}(\mathcal{Q}) \rightarrow L^2(\mathcal{Q}), \quad S_f(u) = \partial_t u - \frac{1}{2} \Delta_x u + fu,$$

where

$$H_{B,0}^{2,1}(\mathcal{Q}) = \{u \in H^{2,1}(\mathcal{Q}) \mid u = 0 \text{ on } \Sigma \text{ and } u(x, 0) = 0\}.$$

Then S_f is an isomorphism with a linear continuous inverse operator

$$V_f : L^2(\mathcal{Q}) \rightarrow H_{B,0}^{2,1}(\mathcal{Q}), \quad h \mapsto V_f(h).$$

That is, for any $h \in L^2(\mathcal{Q})$ the inhomogeneous equation

$$\begin{cases} \partial_t u - \frac{1}{2} \Delta_x u + fu = h & \text{on } \mathcal{Q} \\ u = 0 & \text{on } \Sigma \\ u(\cdot, 0) = 0 & \text{on } \mathcal{O} \end{cases} \quad (31)$$

has a unique weak solution $w_{f,h} = V_f(h) \in H_{B,0}^{2,1}(\mathcal{Q})$, [42, chapter 4, remark 15.1].

Lemma 12. *There exists a constant $C > 0$, such that for all $f \in C(\bar{\mathcal{O}})$, with $f > 0$, and $h : [0, \mathbf{T}] \times \bar{\mathcal{O}} \rightarrow \mathbb{R}$ a continuous function with $t \mapsto h(t, \cdot) \in C([0, \mathbf{T}]; C(\bar{\mathcal{O}}))$ the solution $w_{f,h}$ to (31) satisfies*

$$\|w_{f,h}\|_{L^2(\mathcal{Q})} \leq C \|h\|_{L^2(\mathcal{Q})}. \quad (32)$$

Proof. The solution $w_{f,h}$ to (31) has a presentation

$$w_{f,h}(x, t) = \int_0^t e^{(t-s)S_f} h(\cdot, s) ds(x), \quad 0 \leq t \leq \mathbf{T}, \quad x \in \bar{\Omega},$$

see [43, theorem 5.1.11]. Using Hölder's inequality we can then write

$$\begin{aligned} \|w_{f,h}\|_{L^2}^2 &= \int_Q \left(\int_0^t e^{(t-s)S_f} h(\cdot, s) ds(x) \right)^2 d(x, t) \\ &\leq \int_Q \int_0^t e^{2(t-s)S_f} ds(x) \int_0^t h^2(x, s) ds d(x, t) \\ &= \int_Q \frac{1}{2} \int_0^{2t} e^{(2t-\tilde{s})S_f} d\tilde{s}(x) \int_0^t h^2(x, s) ds d(x, t) \\ &\leq \int_Q \int_0^{\mathbf{T}} \frac{1}{2} |w_{f,1}(x, 2t)| dt \int_0^{\mathbf{T}} h^2(x, s) ds dx \\ &\leq \mathbf{T} \|w_{f,1}\|_{\infty} \|h\|_{L^2}^2. \end{aligned}$$

The solution to (31) has also a probabilistic representation in terms of the Feynman–Kac formula

$$w_{f,h}(x, t) = \mathbb{E}^x \left(\int_0^{\tau_t} h(X_s, t-s) e^{-\int_0^s f(X_r) dr} ds \right),$$

where $(X_s : s \geq 0)$ is a d -dimensional Brownian motion started at $x \in \mathcal{O}$, with exit time $\tau_{\mathcal{O}}$ satisfying $\sup_{x \in \mathcal{O}} \mathbb{E}^x(\tau_{\mathcal{O}}) < \infty$, and $\tau_t = \min\{\tau_{\mathcal{O}}, t\}$. Hence we get a bound

$$\|w_{f,1}\|_{\infty} = \sup_{(x,t) \in Q} \left| \mathbb{E}^x \left(\int_0^{\tau_t} e^{-\int_0^s f(X_r) dr} ds \right) \right| \leq \sup_{(x,t) \in Q} \mathbb{E}^x(\tau_t) \leq \mathbf{T}.$$

□

Lemma 12 can be used to prove the following stronger regularity estimates.

Lemma 13. *Let f, h be as in lemma 12, and $w_{f,h} \in H_{B,0}^{2,1}(Q)$ be the unique solution to (31). Then there exists a constant $C > 0$ such that*

$$\|w_{f,h}\|_{H^{2,1}(Q)} \leq C(1 + \|f\|_{\infty}) \|h\|_{L^2(Q)} \quad (33)$$

$$\|w_{f,h}\|_{L^2(Q)} \leq C(1 + \|f\|_{\infty}) \|h\|_{(H_{C,0}^{2,1}(Q))^*}, \quad (34)$$

where

$$H_{C,0}^{2,1}(Q) = \{u \in H^{2,1}(Q) \mid u = 0 \text{ on } \Sigma \text{ and } u(x, \mathbf{T}) = 0\}.$$

Proof. Using the fact that $S_0 : H_{B,0}^{2,1}(Q) \rightarrow L^2(Q)$ is an isomorphism and lemma 12 we get

$$\begin{aligned} \|w_{f,h}\|_{H^{2,1}} &\leq C \left\| \left(\partial_t - \frac{1}{2} \Delta_x \right) w_{f,h} \right\|_{L^2} \\ &\leq C (\|S_f(w_{f,h})\|_{L^2} + \|f w_{f,h}\|_{L^2}) \\ &\leq C (\|h\|_{L^2} + \|f\|_{\infty} \|w_{f,h}\|_{L^2}) \\ &\leq C(1 + \mathbf{T} \|f\|_{\infty}) \|h\|_{L^2} \end{aligned}$$

which proves the first estimate.

Denote by $S_f^* = -\partial_t - \frac{1}{2}\Delta_x + f$ the adjoint of S_f and by $H_{0,0}^{2r,r}(Q)$ the closure of $C_c^\infty(Q)$ in $H^{2r,r}(Q)$. Let $\varphi \in H_{0,0}^{2(r-1),r-1}(Q)$, with some $r \geq 1$. Then the adjoint problem

$$\begin{cases} S_f^*(u) = \varphi & \text{on } Q \\ u = 0 & \text{on } \Sigma \\ u(\cdot, \mathbf{T}) = 0 & \text{on } \mathcal{O} \end{cases} \tag{35}$$

has a solution $v_{f,\varphi}^* \in X^r(Q)$, where

$$X^r(Q) = \{u \in H^{2r,r}(Q) : u = 0 \text{ on } \Sigma, u(x, \mathbf{T}) = 0 \text{ and } S_f^*(u) \in H_{0,0}^{2(r-1),r-1}(Q)\}.$$

The adjoint operator S_f^* is an isomorphism of $X^r(Q)$ onto $H_{0,0}^{2(r-1),r-1}(Q)$, $r \geq 1$, and we denote the inverse operator by $V_f^* : H_{0,0}^{2(r-1),r-1}(Q) \rightarrow X^r(Q)$, [42, section 7].

We start by showing that the estimate (32) holds also for the solution to the adjoint problem (35). Let $\varphi \in C_c^\infty(Q)$ and $\varphi \neq 0$. Then $v_{f,\varphi}^* \neq 0$, and we get

$$\begin{aligned} \|v_{f,\varphi}^*\|_{L^2}^2 &= \int v_{f,\varphi}^* V_f^*(\varphi) \\ &\leq \|V_f(v_{f,\varphi}^*)\|_{L^2} \|\varphi\|_{L^2} \\ &= \|w_{f,v_{f,\varphi}^*}\|_{L^2} \|\varphi\|_{L^2} \\ &\leq \mathbf{T} \|v_{f,\varphi}^*\|_{L^2} \|\varphi\|_{L^2}, \end{aligned}$$

where the last inequality follows from lemma 12. From the above we see that $\|v_{f,\varphi}^*\|_{L^2} \leq \mathbf{T} \|\varphi\|_{L^2}$.

We can then show, with a similar proof to that of (33), that

$$\|v_{f,\varphi}^*\|_{H^{2,1}} \leq (1 + \mathbf{T}\|f\|_\infty) \|\varphi\|_{L^2},$$

and conclude

$$\begin{aligned} \|w_{f,h}\|_{L^2} &= \sup_{\varphi \in C_c^\infty, \|\varphi\|_{L^2} \leq 1} \left| \int_Q w_{f,h} \varphi \right| \\ &= \sup_{\varphi \in C_c^\infty, \|\varphi\|_{L^2} \leq 1} \left| \int_Q w_{f,h} S_f^*(V_f^*(\varphi)) \right| \\ &= \sup_{\varphi \in C_c^\infty, \|\varphi\|_{L^2} \leq 1} \left| \int_Q S_f(w_{f,h}) V_f^*(\varphi) \right| \\ &\leq \sup_{\varphi \in C_c^\infty, \|\varphi\|_{L^2} \leq 1} \|v_{f,\varphi}^*\|_{H^{2,1}} \|h\|_{(H_{C,0}^{2,1})^*} \\ &\leq (1 + \mathbf{T}\|f\|_\infty) \|h\|_{(H_{C,0}^{2,1})^*}. \end{aligned}$$

□

We now turn to the properties of the forward map G . The following norm estimate for the $\mathcal{C}^{2,1}$ -Hölder–Zygmund norm of $G(f) = u_f$ is needed for the proof of the proposition 9.

Lemma 14. *Let g, u_0 and f be as in section 2.2. Then for u_f , the unique solution to (2), there exists a constant $C > 0$ such that*

$$\|u_f\|_{\mathcal{C}^{2,1}(\mathcal{Q})} \leq C (1 + \|f\|_\infty) (\|g\|_{\mathcal{C}^{2,1}(\Sigma)} + \|u_0\|_{\mathcal{C}^2(\mathcal{O})})$$

Proof. We start by noticing that since $f > 0$

$$\begin{aligned} \|u_f\|_\infty &= \sup_{(x,t) \in \mathcal{Q}} \left| \mathbb{E}^x \left(u_0(X_t) \chi_{\tau_t=t} \exp \left(- \int_0^t f(X_s) ds \right) \right) \right. \\ &\quad \left. + \mathbb{E}^x \left(g(X_{\tau_{\mathcal{O}}}, \tau_{\mathcal{O}}) \chi_{\tau_t < t} \exp \left(- \int_0^{\tau_{\mathcal{O}}} f(X_s) ds \right) \right) \right| \\ &\leq \|u_0\|_\infty + \|g\|_\infty. \end{aligned} \tag{36}$$

We will also need the isomorphism, see e.g. [5],

$$\begin{aligned} \left(\partial_t - \frac{1}{2} \Delta_x, \text{tr}_\Sigma, \text{tr}_{\mathcal{O}} \right) : \mathcal{C}^{2,1}(\mathcal{Q}) &\rightarrow \mathcal{X} \\ u &\mapsto \left(\partial_t u - \frac{1}{2} \Delta_x u, \text{tr}_\Sigma(u), u(\cdot, 0) \right), \end{aligned}$$

where \mathcal{X} is the subspace of $C(\mathcal{Q}) \times \mathcal{C}^{2,1}(\Sigma) \times \mathcal{C}^2(\mathcal{O})$ of the elements (h, g, u_0) satisfying the consistency conditions

$$g(x, 0) = u_0(x) \quad \text{and} \quad \partial_t g(x, 0) - \frac{1}{2} \Delta_x u_0(x) = h(x, 0), \quad x \in \partial \mathcal{O}.$$

Using the above we get

$$\begin{aligned} \|u_f\|_{\mathcal{C}^{2,1}(\mathcal{Q})} &\leq C \left(\left\| \left(\partial_t - \frac{1}{2} \Delta_x \right) u_f \right\|_\infty + \|g\|_{\mathcal{C}^{2,1}(\Sigma)} + \|u_0\|_{\mathcal{C}^2(\mathcal{O})} \right) \\ &= C (\|f u_f\|_\infty + \|g\|_{\mathcal{C}^{2,1}(\Sigma)} + \|u_0\|_{\mathcal{C}^2(\mathcal{O})}) \\ &\leq C (\|f\|_\infty \|u_f\|_\infty + \|g\|_{\mathcal{C}^{2,1}(\Sigma)} + \|u_0\|_{\mathcal{C}^2(\mathcal{O})}) \\ &\leq C (1 + \|f\|_\infty) (\|g\|_{\mathcal{C}^{2,1}(\Sigma)} + \|u_0\|_{\mathcal{C}^2(\mathcal{O})}). \end{aligned}$$

□

Using the above lemmas we can show that the forward operator G satisfies the following Lipschitz condition.

Lemma 15. *Let g, u_0 and f be as in section 2.2. Then, for the unique solution u_f to (2), there exists a constant $C > 0$ such that*

$$\|u_{f_1} - u_{f_2}\|_{L^2(\mathcal{Q})} \leq C(1 + \|f_1\|_\infty)(1 + \|f_2\|_\infty) \|f_1 - f_2\|_{(H_0^2(\mathcal{O}))^*}.$$

Proof. Let u_{f_i} , $i = 1, 2$, be solutions to (2). We notice that $w = u_{f_1} - u_{f_2}$ solves the inhomogeneous equation $S_{f_1}(w) = (f_1 - f_2)u_{f_2}$ on Q and $w = 0$ on the boundary $\Sigma \times \overline{O}$. Using lemmas 13 and 14 we see that

$$\begin{aligned} \|u_{f_1} - u_{f_2}\|_{L^2} &= \|w_{f_1, (f_1 - f_2)u_{f_2}}\|_{L^2} \\ &\leq C(1 + \mathbf{T}\|f_1\|_\infty)\|(f_1 - f_2)u_{f_2}\|_{(H_{C,0}^{2,1})^*} \\ &\leq C(1 + \mathbf{T}\|f_1\|_\infty) \sup_{\varphi \in H_{C,0}^{2,1}, \|\varphi\|_{H^{2,1}} \leq 1} \left| \int_Q (f_1 - f_2)u_{f_2}\varphi \right| \\ &\leq C(1 + \mathbf{T}\|f_1\|_\infty) \sup_{\varphi \in H_{C,0}^{2,1}, \|\varphi\|_{H^{2,1}} \leq 1} \|u_{f_2}\varphi\|_{H^{2,1}} \|f_1 - f_2\|_{(H_{C,0}^{2,1})^*} \\ &\leq C(1 + \mathbf{T}\|f_1\|_\infty)\|u_{f_2}\|_{C^{2,1}} \|f_1 - f_2\|_{(H_{C,0}^{2,1})^*} \\ &\leq C(1 + \|f_1\|_\infty)(1 + \|f_2\|_\infty)\|f_1 - f_2\|_{(H_{C,0}^{2,1})^*}. \end{aligned}$$

Furthermore

$$\begin{aligned} \|f_1 - f_2\|_{(H_{C,0}^{2,1})^*} &= \sup_{\varphi \in H_{C,0}^{2,1}, \|\varphi\|_{H^{2,1}} \leq 1} \left| \int_Q (f_1(x) - f_2(x))\varphi(x, t) dx \right| \\ &= \sup_{\varphi \in H_{C,0}^{2,1}, \|\varphi\|_{H^{2,1}} \leq 1} \left\| \int_0^{\mathbf{T}} \varphi(\cdot, t) dt \right\|_{H^2} \|f_1 - f_2\|_{(H_0^2)^*}, \end{aligned}$$

where

$$\begin{aligned} \left\| \int_0^{\mathbf{T}} \varphi(\cdot, t) dt \right\|_{H^2(O)}^2 &= \sum_{|\alpha| \leq 2} \left\| D_x^\alpha \int_0^{\mathbf{T}} \varphi(\cdot, t) dt \right\|_{L^2(O)}^2 \\ &= \sum_{|\alpha| \leq 2} \int_O \left| \int_0^{\mathbf{T}} D_x^\alpha \varphi(\cdot, t) dt \right|^2 dx \\ &\leq \sum_{|\alpha| \leq 2} \int_O \mathbf{T} \int_0^{\mathbf{T}} |D_x^\alpha \varphi(\cdot, t)|^2 dt dx \\ &\leq \mathbf{T} \int_0^{\mathbf{T}} \|\varphi(\cdot, t)\|_{H^2}^2 dt \\ &\leq \mathbf{T} \|\varphi(\cdot, t)\|_{H^{2,1}}^2 \end{aligned}$$

which concludes the proof. \square

We can finally proceed to prove propositions 9–11.

Proof of proposition 9. Note that if $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ is a regular link function in the sense of definition 1 then there exists $C > 0$ such that for all $F \in L^\infty(\mathcal{O})$

$$\|\Phi \circ F\|_\infty \leq C(1 + \|F\|_\infty).$$

Also, for all $F_1, F_2 \in C^2(\mathcal{O})$ there exists $C > 0$ such that

$$\|\Phi \circ F_1 - \Phi \circ F_2\|_{(H^2)^*} \leq C(1 + \|F_1\|_{C^2}^2 + \|F_2\|_{C^2}^2)\|F_1 - F_2\|_{(H^2)^*}.$$

See [52, lemma 29] for a proof. Using lemma 15 and the above estimates we can then write

$$\begin{aligned} \|\mathcal{G}(F_1) - \mathcal{G}(F_2)\|_{L^2(\mathcal{Q})} &\leq C(1 + \|f_1\|_\infty)(1 + \|f_2\|_\infty)\|f_1 - f_2\|_{(H_0^2(\mathcal{O}))^*} \\ &\leq C(1 + \|F_1\|_\infty^2 \vee \|F_2\|_\infty^2)(1 + \|F_1\|_{C^2}^2 \vee \|F_2\|_{C^2}^2) \\ &\quad \times \|F_1 - F_2\|_{(H^2(\mathcal{O}))^*} \\ &\leq C(1 + \|F_1\|_{C^2(\mathcal{O})}^4 \vee \|F_2\|_{C^2(\mathcal{O})}^4)\|F_1 - F_2\|_{(H^2(\mathcal{O}))^*}. \end{aligned}$$

□

Proof of proposition 10. Applying Jensen's inequality we see that

$$\begin{aligned} \inf_{(x,t) \in \mathcal{Q}} u_f(x,t) &= \inf_{(x,t) \in \mathcal{Q}} \left(\mathbb{E}^x \left(u_0(X_t) \chi_{\tau_t=t} e^{-\int_0^t f(X_s) ds} \right) \right. \\ &\quad \left. + \mathbb{E}^x \left(g(X_{\tau_{\mathcal{O}}}, \tau_{\mathcal{O}}) \chi_{\tau_t < t} e^{-\int_0^{\tau_{\mathcal{O}}} f(X_s) ds} \right) \right) \\ &\geq u_{0,\min} e^{-\mathbf{T}\|f\|_\infty} + g_{\min} \inf_{x \in \mathcal{O}} e^{-\|f\|_\infty \mathbb{E}^x(\tau_{\mathcal{O}})} \\ &\geq (u_{0,\min} + g_{\min}) e^{-C_{\mathbf{T}}\|f\|_\infty} > 0, \end{aligned}$$

where $C_{\mathbf{T}} = \max\{\mathbf{T}, \mathbb{E}^x(\tau_{\mathcal{O}})\}$.

Since $f \geq f_{\min} > 0$ the solution u_f is positive and we can write $f(x,t) = \frac{(\frac{1}{2}\Delta_x - \partial_t)u_f(x,t)}{u_f(x,t)}$. Note that $f(x,t)$ is a constant in t . We can then write

$$\begin{aligned} \|f_1 - f_2\|_{L^2(\mathcal{Q})} &= \left\| \frac{(\frac{1}{2}\Delta_x - \partial_t)u_{f_1}}{u_{f_1}} - \frac{(\frac{1}{2}\Delta_x - \partial_t)u_{f_2}}{u_{f_2}} \right\|_{L^2(\mathcal{Q})} \\ &\leq \left\| \frac{(\frac{1}{2}\Delta_x - \partial_t)(u_{f_1} - u_{f_2})}{u_{f_1}} \right\|_{L^2(\mathcal{Q})} + \left\| (u_{f_1}^{-1} - u_{f_2}^{-1}) \left(\frac{1}{2}\Delta_x - \partial_t \right) u_{f_2} \right\|_{L^2} \\ &\lesssim \left(\inf_{(x,t) \in \mathcal{Q}} |u_{f_1}(x,t)| \right)^{-1} \|u_{f_1} - u_{f_2}\|_{H^{2,1}(\mathcal{Q})} + \|u_{f_1}^{-1} - u_{f_2}^{-1}\|_{L^2(\mathcal{Q})} \|f_2 u_{f_2}\|_{C^0}. \end{aligned}$$

To bound the last term we note that

$$\left| \frac{1}{u_{f_1}} - \frac{1}{u_{f_2}} \right| \leq \left| \frac{u_{f_1} - u_{f_2}}{\min\{u_{f_1}^2, u_{f_2}^2\}} \right| \leq (u_{0,\min} + g_{\min})^{-2} e^{2C_{\mathbf{T}}\|f_1 \vee f_2\|_\infty} |u_{f_1} - u_{f_2}|.$$

Combining the above with (36) we get

$$\begin{aligned} \|f_1 - f_2\|_{L^2} &\leq (u_{0,\min} + g_{\min})^{-1} e^{C\mathbf{r}\|f_1\|_\infty} \|u_{f_1} - u_{f_2}\|_{H^{2,1}} \\ &\quad + \|f_2\|_\infty (\|u_0\|_\infty + \|g\|_\infty) (u_{0,\min} + g_{\min})^{-2} e^{C\mathbf{r}\|f_1 \vee f_2\|_\infty} \|u_{f_1} - u_{f_2}\|_{L^2}. \end{aligned}$$

We conclude the proof by noting that $\|f_1 - f_2\|_{L^2(\mathcal{Q})} = \mathbf{T}\|f_1 - f_2\|_{L^2(\mathcal{O})}$. \square

Proof of proposition 11. We start by noting that since $f \in C^\beta \subset H^\beta$ and the assumption on g, u_0 the solution $u_f = G(f) \in H^{2+\beta, 1+\beta/2}$, see [42, theorem 5.3]. Denote by

$$\mathcal{F}_{(\text{C.C.})}^\beta \subset \mathcal{F}^\beta = H^{\beta, \beta/2}(\mathcal{Q}) \times H^{3/2+\beta, 3/4+\beta/2}(\Sigma) \times H^{1+\beta}(\mathcal{O})$$

the subspace of \mathcal{F}^β of elements (h, g, u_0) that satisfy the following compatibility conditions (C.C.); there exists $\psi \in H^{2+\beta, 1+\beta/2}(\mathcal{Q})$ such that

$$\begin{aligned} \psi &= g \quad \text{on } \Sigma, \quad \psi(x, 0) = u_0 \quad \text{on } \mathcal{O} \quad \text{and} \\ \partial_t^k \left(\left(\partial_t - \frac{1}{2} \Delta_x + f \right) \psi \right) \Big|_{t=0} &= \partial_t^k h(x, 0) \quad \text{for } 0 \leq k < \frac{\beta}{2} - \frac{1}{2}. \end{aligned}$$

Using the isomorphism [42, theorem 6.2]

$$\begin{aligned} \left(\partial_t - \frac{1}{2} \Delta_x, \text{tr}_\Sigma, \text{tr}_\mathcal{O} \right) : H^{2+\beta, 1+\beta/2} &\rightarrow \mathcal{F}_{(\text{C.R})}^\beta \\ u &\mapsto \left(\partial_t u - \frac{1}{2} \Delta_x u, \text{tr}_\Sigma(u), u(\cdot, 0) \right), \end{aligned}$$

inequality (4), and the interpolation inequality (24) we get

$$\begin{aligned} \|u_f\|_{H^{2+\beta, 1+\beta/2}} &\lesssim \left\| \left(\partial_t - \frac{1}{2} \Delta_x \right) u_f \right\|_{H^{\beta, \beta/2}} + \|\text{tr}_\Sigma(u_f)\|_{H^{3/2+\beta, 3/4+\beta/2}} + \|u(\cdot, 0)\|_{H^{1+\beta}} \\ &\lesssim \|f u_f\|_{H^{\beta, \beta/2}} + \|g\|_{C^{2+\beta, 1+\beta/2}} + \|u_0\|_{C^{2+\beta}} \\ &\lesssim 1 + \|f\|_{C^\beta} \|u_f\|_{H^{\beta, \beta/2}} \\ &\lesssim 1 + \|f\|_{C^\beta} \|u_f\|_{H^{2+\beta, 1+\beta/2}}^{\frac{\beta}{2+\beta}} \|u_f\|_{L^2}^{\frac{2}{2+\beta}}. \end{aligned}$$

If $\|u_f\|_{H^{2+\beta, 1+\beta/2}} \geq 1$ we can divide both sides by $\|u_f\|_{H^{2+\beta, 1+\beta/2}}^{\frac{\beta}{2+\beta}}$ and otherwise estimate the norm on the right-hand side by 1. Using the second part of proposition 9 we then see that

$$\|u_f\|_{H^{2+\beta, 1+\beta/2}} \lesssim 1 + \|f\|_{C^\beta}^{1+\beta/2} \|u_f\|_{L^2} \lesssim 1 + \|f\|_{C^\beta}^{1+\beta/2} \|u_f\|_\infty \lesssim 1 + \|f\|_{C^\beta}^{1+\beta/2}.$$

\square

Acknowledgments

The author would like to thank Richard Nickl for valuable discussions.

Data availability statement

No new data were created or analysed in this study.

ORCID iDs

Hanne Kekkonen  <https://orcid.org/0000-0003-4700-1724>

References

- [1] Abraham K and Nickl R 2020 On statistical Calderón problems *Math. Stat. Learn.* **2** 165–216
- [2] Agapiou S, Larsson S and Stuart A M 2013 Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems *Stoch. Process. Appl.* **123** 3828–60
- [3] Altmeyer R and Reiss M 2019 Nonparametric estimation for linear SPDEs from local measurements (arXiv:1903.06984)
- [4] Amann H 1995 *Linear and Quasilinear Parabolic Problems (Function spaces vol 2)* (Basel: Birkhäuser)
- [5] Amann H 2010 *Anisotropic Function Spaces and Maximal Regularity for Parabolic Problems Part I vol 6* (Prague: Matfyzpress)
- [6] Arridge S, Maass P, Öktem O and Schönlieb C-B 2019 Solving inverse problems using data-driven models *Acta Numer.* **28** 1–174
- [7] Bachmayr M and Burger M 2009 Iterative total variation schemes for nonlinear inverse problems *Inverse Problems* **25** 105004
- [8] Bal G 2012 Hybrid inverse problems and internal functionals *Inverse Problems and Applications: Inside Out II vol 60* pp 325–68
- [9] Bal G and Uhlmann G 2010 Inverse diffusion theory of photoacoustics *Inverse Problems* **26** 085010
- [10] Benning M, Betcke M M, Ehrhardt M J and Schönlieb C-B 2017 Choose your path wisely: gradient descent in a Bregman distance framework (arXiv:1712.04045)
- [11] Benning M and Burger M 2018 Modern regularization methods for inverse problems *Acta Numer.* **27** 1–111
- [12] Beskos A, Girolami M, Lan S, Farrell P E and Stuart A M 2017 Geometric MCMC for infinite-dimensional inverse problems *J. Comput. Phys.* **335** 327–51
- [13] Bissantz N, Hohage T, Munk A and Ruyngaert F 2007 Convergence rates of general regularization methods for statistical inverse problems and applications *SIAM J. Numer. Anal.* **45** 2610–36
- [14] Bohr J and Nickl R 2021 On log-concave approximations of high-dimensional posterior measures and stability properties in non-linear inverse problems (arXiv:2105.07835)
- [15] Calderón A P 1980 On an inverse boundary value problem *Seminar on Numerical Analysis and Its Applications to Continuum Physics* (Rio de Janeiro) pp 65–73
- [16] Colton D and Kress R 1998 *Inverse Acoustic and Electromagnetic Scattering Theory* 2nd edn (Berlin: Springer)
- [17] Cotter S L, Roberts G O, Stuart A M and White D 2013 MCMC methods for functions: modifying old algorithms to make them faster *Stat. Sci.* **28** 424–46
- [18] Dashti M and Stuart A M 2017 The Bayesian approach to inverse problems *Handbook of Uncertainty Quantification* (Berlin: Springer) pp 311–428
- [19] Daubechies I 2006 *Ten Lectures on Wavelets (Ninth Printing, 2006) (CBMS-NSF Regional Conference Series in Applied Mathematics vol 61)* (Philadelphia, PA: SIAM)
- [20] Dosso S E, Holland C W and Sambridge M 2012 Parallel tempering for strongly nonlinear geoaoustic inversion *J. Acoust. Soc. Am.* **132** 3030–40
- [21] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems (Mathematics and Its Applications vol 375)* (Dordrecht: Kluwer)
- [22] Freidlin M I 1985 *Functional Integration and Partial Differential Equations vol 109* (Princeton, NJ: Princeton University Press)
- [23] Friedman A 1964 *Partial Differential Equations of Parabolic Type* (Englewood Cliffs, NJ: Prentice-Hall)

- [24] Ghosal S and van der Vaart A 2017 *Fundamentals of Nonparametric Bayesian Inference* vol 44 (Cambridge: Cambridge University Press)
- [25] Giné E and Nickl R 2015 *Mathematical Foundations of Infinite-Dimensional Statistical Models* vol 40 (Cambridge: Cambridge University Press)
- [26] Giordano M and Kekkonen H 2020 Bernstein–von Mises theorems and uncertainty quantification for linear inverse problems *SIAM/ASA J. Uncertain. Quantification* **8** 342–73
- [27] Giordano M and Nickl R 2020 Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem *Inverse Problems* **36** 085001
- [28] Hairer M, Stuart A M and Vollmer S J 2014 Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions *Ann. Appl. Probab.* **24** 2455–90
- [29] Hildebrandt F and Trabs M 2019 Parameter estimation for SPDEs based on discrete observations in time and space (arXiv:1910.01004)
- [30] Hohage T and Weidling F 2015 Verification of a variational source condition for acoustic inverse medium scattering problems *Inverse Problems* **31** 075006
- [31] Isaacson D, Mueller J L, Newell J C and Siltanen S 2004 Reconstructions of chest phantoms by the D-bar method for electrical impedance tomography *IEEE Trans. Med. Imaging* **23** 821–8
- [32] Kaipio J and Somersalo E 2005 *Statistical and Computational Inverse Problems (Applied Mathematical Sciences* vol 160) (New York: Springer)
- [33] Kaltenbacher B, Neubauer A and Scherzer O 2008 *Iterative Regularization Methods for Nonlinear Ill-Posed Problems* vol 6 (New York: de Gruyter & Co)
- [34] Kaltenbacher B, Nguyen T T N and Scherzer O 2021 The tangential cone condition for some coefficient identification model problems in parabolic PDEs *Time-Dependent Problems in Imaging and Parameter Identification* pp 121–63
- [35] Kekkonen H, Lassas M and Siltanen S 2016 Posterior consistency and convergence rates for Bayesian inversion with hypoelliptic operators *Inverse Problems* **32** 085005
- [36] Knapik B and Salomond J-B 2018 A general approach to posterior contraction in nonparametric inverse problems *Bernoulli* **24** 2091–121
- [37] Knapik B T, van der Vaart A W and van Zanten J H 2011 Bayesian inverse problems with Gaussian priors *Ann. Stat.* **39** 2626–57
- [38] Krylov N V 1996 *Lectures on Elliptic and Parabolic Equations in Hölder Spaces* vol 12 (Providence, RI: American Mathematical Society)
- [39] Kuchment P and Kunyansky L 2015 Mathematics of photoacoustic and thermoacoustic tomography *Handbook of Mathematical Methods in Imaging* pp 1117–67
- [40] Latz J, Madrigal-Cianci J P, Nobile F and Tempone R 2020 Generalized parallel tempering on Bayesian inverse problems (arXiv:2003.03341)
- [41] Lions J-L and Magenes E 1972 *Non-Homogeneous Boundary Value Problems and Applications* vol 1 (New York: Springer)
- [42] Lions J-L and Magenes E 1972 *Non-Homogeneous Boundary Value Problems and Applications* vol 2 (New York: Springer)
- [43] Lunardi A 2012 *Analytic Semigroups and Optimal Regularity in Parabolic Problems* (Berlin: Springer)
- [44] Ma Y-A, Chen T and Fox E B 2015 A complete recipe for stochastic gradient MCMC *Proc. 28th Int. Conf. Neural Information Processing Systems* vol 2 pp 2917–25
- [45] Meyer Y 1995 *Wavelets and Operators* vol 1 (Cambridge: Cambridge University Press)
- [46] Monard F, Nickl R and Paternain G P 2019 Efficient nonparametric Bayesian inference for x-ray transforms *Ann. Stat.* **47** 1113–47
- [47] Monard F, Nickl R and Paternain G P 2020 Statistical guarantees for Bayesian uncertainty quantification in non-linear inverse problems with Gaussian process priors (arXiv:2007.15892)
- [48] Monard F, Nickl R and Paternain G P 2021 Consistent inversion of noisy non-abelian x-ray transforms *Commun. Pure Appl. Math.* **74** 1045–99
- [49] Nickl R 2020 Bernstein–von Mises theorems for statistical inverse problems: I. Schrödinger equation *J. Eur. Math. Soc.* **22** 2697–750
- [50] Nickl R and Söhl J 2017 Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions *Ann. Stat.* **45** 1664–93
- [51] Nickl R and Söhl J 2019 Bernstein–von Mises theorems for statistical inverse problems: II. Compound Poisson processes *Electron. J. Stat.* **13** 3513–71
- [52] Nickl R, van de Geer S and Wang S 2020 Convergence rates for penalized least squares estimators in PDE constrained regression problems *SIAM/ASA J. Uncertain. Quantification* **8** 374–413

- [53] Nickl R and Wang S 2020 On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms (arXiv:2009.05298)
- [54] Patriarca M and Leppänen T 2004 Modeling language competition *Physica A* **338** 296–9
- [55] Ray K 2013 Bayesian inverse problems with non-conjugate priors *Electron. J. Stat.* **7** 2516–49
- [56] Solé R V, Corominas-Murtra B and Fortuny J 2010 Diversity, competition, extinction: the eco-physics of language change *J. R. Soc. Interface* **7** 1647–64
- [57] Stuart A M 2010 Inverse problems: a Bayesian perspective *Acta Numer.* **19** 451–559
- [58] Szabó B, van der Vaart A and van Zanten J 2015 Frequentist coverage of adaptive nonparametric Bayesian credible sets *Ann. Stat.* **43** 1391–428
- [59] Triebel H 1983 *Theory of Function Spaces (Monographs in Mathematics vol 78)* (Basel: Birkhäuser)
- [60] Triebel H 2008 *Function Spaces and Wavelets on Domains (Tracts in Mathematics vol 7)* (Zürich: European Mathematical Society)
- [61] Vollmer S J 2013 Posterior consistency for Bayesian inverse problems through stability and regression results *Inverse Problems* **29** 125011
- [62] Wu J 2012 *Theory and Applications of Partial Functional Differential Equations vol 119* (Berlin: Springer)