Submitted to the Annals of Statistics

LINEAR METHODS FOR NON-LINEAR INVERSE PROBLEMS

BY GEERTEN KOERS[1], BOTOND SZABÓ[2] AND AAD VAN DER VAART[1]

[1]*Delft University of Technology, DIAM, a.w.vandervaart@tudelft.nl*

[2]*Bocconi University, Department of Data Sciences and BIDSA, botond.szabo@unibocconi.it*

We consider the recovery of an unknown function f from a noisy observation of the solution u_f to a partial differential equation that can be written in the form $\mathcal{L}u_f = c(f, u_f)$, for a differential operator $\mathcal{L}$ that is rich enough to recover f from $\mathcal{L}u_f$. Examples include the time-independent Schrödinger equation $\Delta u_f = 2u_f f$, the heat equation with absorption term $(\partial_t - \Delta_x/2)u_f = f u_f$, and the Darcy problem $\nabla \cdot (f \nabla u_f) = h$. We transform this problem into the linear inverse problem of recovering $\mathcal{L}u_f$ under the Dirichlet boundary condition, and show that Bayesian methods with priors placed either on u_f or $\mathcal{L}u_f$ for this problem yield optimal recovery rates not only for u_f, but also for f. We also derive frequentist coverage guarantees for the corresponding Bayesian credible sets. Adaptive priors are shown to yield adaptive contraction rates for f, thus eliminating the need to know the smoothness of this function. The results are illustrated by numerical experiments on synthetic data sets.

1. Introduction.

Partial differential equation (PDE) constrained statistical inverse problems arise naturally in various fields of sciences, including medical imaging, astronomy, physics and the earth sciences. In this paper we consider the recovery of various parameters of an equation, e.g., the diffusivity or absorption terms, from a noisy observation of the solution, using nonparametric Bayesian methods.

As a concrete example of a PDE, consider Darcy's problem in the context of groundwater flow. For a bounded domain $\mathcal{O} \subset \mathbb{R}^d$, an unknown function $f : \mathcal{O} \to \mathbb{R}$ modelling the diffusivity (or conductivity) of the medium $\mathcal{O} \subset \mathbb{R}^d$ and given functions $h : \mathcal{O} \to \mathbb{R}$ and $g : \partial\mathcal{O} \to \mathbb{R}$, consider the "steady state" solution $u_f : \bar{\mathcal{O}} \to \mathbb{R}$ of the elliptic PDE

$$\begin{cases} \operatorname{div}(f \nabla u_f) = h, & \text{on } \mathcal{O}, \\ u_f = g, & \text{on } \partial\mathcal{O}. \end{cases} \tag{1.1}$$

Here $\operatorname{div} v = \nabla \cdot v$ denotes the divergence $\operatorname{div} v = \sum_{i=1}^d \frac{\partial}{\partial x_i} v_i$ of a vector field $v : \mathcal{O} \to \mathbb{R}^d$ and ∇u is the gradient of a function $u : \mathcal{O} \to \mathbb{R}$, so that $\operatorname{div}(f \nabla u) = \nabla f \cdot \nabla u + f \Delta u$, with the dot on the right side the ordinary inner product of $\mathbb{R}^d$ and Δ the Laplacian. We are interested in recovering f from a noisy observation of u_f, given known functions g and h.

As a second example, consider the heat equation with an absorption term. For a given (smooth) absorption function $f \in L_2\big((0,1)^d \times (0,1)\big)$, an initial value function $u_0 \in L_2((0,1)^d)$ and a boundary condition $g \in L_2\big(\partial(0,1)^d \times [0,1]\big)$, we consider the solution

MSC2020 subject classifications: Primary 62G05, 62G15; secondary 62G20.

Keywords and phrases: Posterior distribution, Inverse problem, non-linear, Contraction rate, Adaptation, Uncertainty quantification.

$u_f : [0,1]^d \times [0,1] \to \mathbb{R}$ to the parabolic partial differential equation

$$(1.2)\qquad \begin{cases} \dfrac{\partial}{\partial t}u - \dfrac{1}{2}\Delta u = fu, & \text{on } (0,1)^d \times (0,1], \\ u = g, & \text{on } \partial(0,1)^d \times [0,1], \\ u = u_0, & \text{on } (0,1)^d \times \{0\}. \end{cases}$$

Again inference is on f given a noisy observation of u_f, and given known functions g, u_0.

In both cases, we take the noise to be additive and Gaussian. We write the observation informally as

$$(1.3)\qquad Y_n = u_f + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}.$$

We consider two precise versions of this model. In the white noise model the observation is the Gaussian process $\big(Y_n(h) : h \in L_2(\mathcal{O})\big)$, given by

$$Y_n(h) = \langle h, u_f\rangle_{L_2} + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}(h),$$

for $\dot{\mathbb{W}} = \big(\dot{\mathbb{W}}(h) : h \in L_2(\mathcal{O})\big)$ the mean zero Gaussian process with covariance function $\mathrm{cov}\big(\dot{\mathbb{W}}(h_1), \dot{\mathbb{W}}(h_2)\big) = \langle h_1, h_2\rangle_{L_2}$. Next to this continuous interpretation of (1.3), we also consider the practically relevant discrete observational setting, where Y_n is observed at n given design points $x_{n,1}, \ldots, x_{n,n} \in \mathcal{O}$. In this case the observation Y_n is a vector in $\mathbb{R}^n$ with coordinates $Y_n(i) = u_f(x_{n,i}) + W_i$, where $W_1, \ldots, W_n$ are i.i.d. standard normal variables. In our general discussion we shall use (1.3) to refer to both models.

The paper addresses a more general formulation of PDE-constrained statistical inverse problems, which includes Darcy's problem and the heat equation with absorption as examples. We consider the recovery of a function $f : \mathcal{O} \to \mathbb{R}$ on a known domain $\mathcal{O} \subset \mathbb{R}^d$ from a noisy observation of the function u_f satisfying a PDE of the form

$$(1.4)\qquad \begin{cases} \mathcal{L}u_f = c(f, u_f), & \text{on } \mathcal{O}, \\ u_f = g, & \text{on } \Gamma \subseteq \partial\mathcal{O}. \end{cases}$$

Here $\mathcal{L}$ is a linear differential operator and $g : \Gamma \to \mathbb{R}$ is a known function on a subset Γ of the boundary $\partial\mathcal{O}$ of $\mathcal{O}$. The map c is known and transforms the pair of functions (f, u_f) into a function on $\mathcal{O}$. In most of our examples this map takes the form of a pointwise transformation, such as $c(f, u_f)(x) = \bar{c}\big(f(x), u_f(x)\big)$, for a map $\bar{c} : \mathbb{R}^2 \to \mathbb{R}$, but c may also act on (f, u_f) as elements of given function spaces (and hence also involve derivatives of f or u_f).

A PDE can typically be presented in the form (1.4) in various ways. For the Darcy problem (1.1), we shall choose $\mathcal{L} = \Delta$ and $c(f, u_f) = (h - \nabla f \cdot \nabla u)/f$, while for the heat equation model (1.2), we take $\mathcal{L} = \frac{\partial}{\partial t} - \frac{1}{2}\Delta$ and $c(f, u_f) = f u_f$. These choices are motivated by the idea to choose the linear operator $\mathcal{L}$ rich enough to be able to recover the function f from $\mathcal{L}u_f$, and to isolate possible non-linearities in the solution map $u_f \mapsto f$ of the PDE in the term $c(f, u_f)$. We make the first precise by assuming that there is a (sufficiently regular) solution map e such that

$$(1.5)\qquad f = e(\mathcal{L}u_f).$$

In several examples, the map e takes a simple, pointwise form, which makes our general method very easy to implement. For instance, in case of the heat equation $e(v) = v/(\mathcal{L}^{-1}v)$. More generally, the map e may act on $\mathcal{L}u_f$ as an element of a function space, and practical implementation may involve a numerical implementation of the recovery of f from $\mathcal{L}u_f$, as is true for Darcy's problem.

The general idea of the paper is to solve the (linear) inverse problem of recovering $\mathcal{L}u_f$ from the observations by a linear Bayesian method, and next recover f by an implementation of (1.5). We prove that the method can attain optimal results, and confirm the feasibility of the approach also by numerical experiments. The two-step approach has both computational and theoretical advantages.

The two examples in the preceding, as well as additional ones, are discussed in Sections 4–9, starting with a detailed discussion of our general method on one of the simplest PDE models, the time-independent Schrödinger equation, $\frac{1}{2}\Delta u_f = u_f f$, considered in [25, 27, 29], where $\mathcal{L} = \Delta$ the Laplacian, $c(u, f) = 2uf$ and $e(v) = v/(2\mathcal{L}^{-1}v)$, in Section 4. Following this, we consider the heat equation with an absorption term, Darcy's problem in one dimension, and subsequently in general dimension d, and finally the exponential of the Voltera operator. It is not clear that the approach works and is useful in general, but this list of examples is encouraging. The approach connects to existing approaches for other inverse problems (e.g. [14]) and may be useful for a Bayesian analysis of such problems as well. A non-homogeneous boundary condition ($u_f = g$ on Γ with g nonzero) still renders the problem of recovering $\mathcal{L}u_f$ from u_f non-linear. To remove also this non-linearity, define K as the solution operator $u \mapsto Ku$ of the homogeneous partial differential equation

$$\begin{cases} \mathcal{L}Ku = u, & \text{on } \mathcal{O}, \\ Ku = 0, & \text{on } \Gamma \subseteq \partial\mathcal{O}. \end{cases} \tag{1.6}$$

Thus the linear operator K is the inverse of $\mathcal{L}$ under the Dirichlet boundary condition. (For our purposes it suffices that K is defined and satisfies (1.6) on the smaller set of functions $\{\mathcal{L}u : u = 0 \text{ on } \Gamma\}$.) Furthermore, for the given boundary function $g : \Gamma \to \mathbb{R}$ in (1.4), let $\tilde{g} : \mathcal{O} \to \mathbb{R}$ be the function that satisfies

$$\begin{cases} \mathcal{L}\tilde{g} = 0, & \text{on } \mathcal{O}, \\ \tilde{g} = g, & \text{on } \Gamma \subseteq \partial\mathcal{O}. \end{cases} \tag{1.7}$$

Solutions to the equations (1.6)-(1.7) exist and are unique under mild regularity conditions on the operator $\mathcal{L}$, domain $\mathcal{O}$, boundary set Γ and function g. In particular, suppose that the solution to the problem to (1.7) with zero boundary condition is unique (i.e. the function $v = 0$ is the only solution to the problem $\mathcal{L}v = 0$ on $\mathcal{O}$ and $v = 0$ on Γ). Because the definitions (1.6) and (1.7) imply that $\mathcal{L}(K\mathcal{L}u_f + \tilde{g}) = \mathcal{L}K(\mathcal{L}u_f) + 0 = \mathcal{L}u_f$ on $\mathcal{O}$, and $K\mathcal{L}u_f + \tilde{g} = 0 + g$ on Γ, it then follows that

$$u_f = K\mathcal{L}u_f + \tilde{g}. \tag{1.8}$$

Since the function $\tilde{g}$ is known, this and (1.3) leads to the adapted observational model

$$\tilde{Y}_n := Y_n - \tilde{g} = K(\mathcal{L}u_f) + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}. \tag{1.9}$$

Thus the problem of estimating $\mathcal{L}u_f$ has been reduced to solving the linear inverse problem $\tilde{Y}_n = Kv + n^{-1/2}\dot{\mathbb{W}}$, defined by the operator K. The solution v to this problem is next substituted as $v = \mathcal{L}u_f$ in the inversion map (1.5) to find f. In the Bayesian approach, the induced distribution of f when this map is applied to a posterior distribution of v is the usual posterior distribution of f for the given prior. (We give a precise statement in Section 2.)

It is natural to put a prior distribution directly on the function f, which is the basic parameter. Within the context of non-linear inverse problems this approach is taken in [17, 20, 25–28], who put a Gaussian process prior on f or $\log f$ or independent priors on the coefficients of an orthonormal wavelet basis of the domain. However, in our two-step approach, which uses the Bayesian method to recover $\mathcal{L}u_f$ in the observational model (1.9), it is easier to put a prior on $\mathcal{L}u_f$ or u_f. Because f is determined uniquely by $\mathcal{L}u_f$ in view of (1.5), this induces a prior (and posterior) distribution on f. Methods for computing the posterior distribution of $\mathcal{L}u_f$ given a prior on this function, as well as theoretical results on this posterior distribution are available (e.g. [1, 18, 22–24, 33, 53, 54]).

If the prior distribution puts probability one on functions $v = \mathcal{L}u_f$ in the domain of the solution map e, then so does the posterior distribution, and the final output of our two-step method is an ordinary Bayesian analysis for the function f. To ease computation, we may also allow prior distributions that give positive probability to functions v not of the form $\mathcal{L}u_f$, for some f in the target model. Under our conditions, the resulting posterior distribution of v will with probability tending to one (typically at exponential speed) concentrate on the domain of e and hence induce a posterior distribution on f through (1.5). This happens, because we choose the operator $\mathcal{L}$ rich enough, so that the posterior distribution contracts in suitable, strong norms for $\mathcal{L}u_f$ and u_f. Thus even if we choose an easy prior on v, which may ignore the underlying PDE, the posterior distribution eventually concentrates on a set where the full information of the PDE can be exploited, through the solution map e. We show that nothing is lost at the level of contraction rates, and our numerical experiments also show satisfactory performance. We note that the procedure is different from projecting an ordinary nonparametric regression estimate (or posterior) (using a weak norm) on the surface of functions u_f generated by the PDE.

The posterior distribution is the conditional distribution given the data Y_n, as usual, where the distribution of the data given the parameter is determined by (1.3). it is denoted by $\Pi_n(\cdot|\,Y_n)$, where the argument $\cdot$ can be specialised to a set concerning u_f, $\mathcal{L}u_f$ or f.

We obtain posterior contraction rates, as usual ([15]), in the non-Bayesian setup where the observation is assumed to be generated according to the model (1.4) and (1.3) with a fixed parameter f_0. For instance, a contraction rate for f relative to a given norm $\|\cdot\|$ is a sequence $\epsilon_n \to 0$ such that $\Pi_n\big(f : \|f - f_0\| \geq M_n\epsilon_n |\, Y_n\big) \to 0$ in probability, for every $M_n \to \infty$. This rate can be compared to an optimal contraction rate, which is determined, in a minimax sense, by the smoothness of the function f_0. The optimal contraction rate is typically obtainable by using a prior that matches the smoothness of the true parameters. By mixing over priors of different smoothness levels (hierarchical Bayes) or using a prior of data-determined smoothness level (empirical Bayes), optimal contraction rates for a range of different smoothness levels are obtainable by a single prior. An advantage of our approach is that such "adaptive" contraction rates for the linear inverse problem of estimating $\mathcal{L}u_f$ carry over into adaptive rates for the recovery of f. Adaptive priors for the linear problem have been constructed by using a scale of priors of fixed smoothness simultaneously through empirical or hierarchical Bayes methods (e.g. [2, 18, 22, 39, 45]).

Besides for estimating the parameter at a (nearly) minimax rate, a posterior distribution can be used for uncertainty quantification. This can be justified, or not, from a non-Bayesian point of view by the coverage of credible sets, i.e. the probability that a (central) set of given posterior probability covers the true parameter f_0. For infinite-dimensional credible sets that involve a bias-variance trade-off, such as balls and bands, this was studied for linear inverse problems in [22, 38, 42, 44]. Coverage is different for priors of fixed smoothness and adaptive priors, and depends on properties of the true parameter relative to the prior. Since our method is based on transforming the non-linear problem at hand into a linear one, the positive results showing the existence of reasonable credible sets in the linear problem imply such results for the non-linear problem of interest (see Section 2). Coverage of credible intervals for smooth functionals of the parameter or of credible balls in a weak norm follow from a suitable version of the Bernstein-von Mises theorem ([6, 7, 25, 27, 34]). The question whether such results are valid for the main construction in the present paper is currently under study.

The approach of this paper is partly motivated by computational efficiency. The recent paper [30] derived a Langevin type algorithm to compute the posterior mean, which provenly takes only polynomial time. The algorithm is based on only the forward map $f \mapsto u_f$, which is an advantage for many problems. However, repeatedly evaluating the forward map may still be heavy for some practical applications in case of medium or large data sets. A gain of our approach is that the linear problem, to compute $\mathcal{L}u_f$, may be substantially faster to solve than the original non-linear one. If the inverse map (1.5) is easy (as it is in the case of most of our examples), then this advantage is propagated to the computation of f. A further gain is that our approach may combine well with distributed computation based on partitioning the data in space. Since the observation is a noisy version of the function $u_f : \mathcal{O} \to \mathbb{R}$, a (possibly overlapping) partition of the domain $\mathcal{O}$, followed by separate reconstruction of $\mathcal{L}u_f$ on the partitioning sets, is feasible. Preliminary work shows that such a distributional approach may give theoretically optimal and practically competitive results compared to other computational shortcuts. In particular, spatial partitioning allows priors and posteriors to adapt to the smoothness of the true function (for the regression problem without inversion, see [43]). In comparison, an approach based on partitioning the domain of the function of f seems impossible, due to the structure of the partial differential equation. Another approach to speed up the computations is the extended version of the inducing variable variational Bayes approach [47]. By optimally choosing the number of inducing variables, this procedure recovers the functional parameter of interest with the optimal minimax rate [32] and (in case of the direct regression model) provides reliable uncertainty quantification [31, 48].

The paper is organised as follows. In Section 2 we formalise the relation between the non-linear inverse problem of interest and the auxiliary linear inverse problem. Next in Section 3 we review and extend Bayesian methods for the linear inverse problem, describing prior choices, with both fixed and data driven choices of the tuning parameter, and resulting results on contraction rates and frequentist coverage of credible sets. In Sections 4–9, we apply our approach to four examples of PDEs. We conclude with a discussion of the approach and open questions in Section 11. In Section 10 of the Supplementary Material we present numerical simulations that illustrate our approach in various settings. Sections A–E provide technical complements. Sections F and G presents posterior contraction rates in smoothness norms for the linear inverse problem in the Gaussian white noise, and the discrete observation settings, respectively.

1.1. *Notation.* For two sequences a_n, b_n we write $a_n \lesssim b_n$ if there exists a constant $C > 0$ such that $a_n / b_n \leq C$, for every n. We write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold simultaneously. The letters c and C denote constants not depending on n; their values might change from line to line.

A *domain* is an open set $\mathcal{O} \subset \mathbb{R}^d$, which throughout the paper we shall take to be bounded and connected with a Lipschitz boundary.

For $\beta \in \mathbb{R}$ the notation $H^\beta(\mathcal{O})$ is the Sobolev space $W^{\beta,2}(\mathcal{O})$ of functions (or distributions) $f : \mathcal{O} \to \mathbb{R}$ of smoothness β. For $\beta = 0$ this is $L_2(\mathcal{O})$, while for $\beta \in \mathbb{N}$ this consists of the functions $f : \mathcal{O} \to \mathbb{R}$ with weak derivatives $D^\alpha f$ up to order $|\alpha| \leq \beta$, normed by

$$\|f\|_{H^\beta(\mathcal{O})} := \sqrt{\sum_{0 \leq |\alpha| \leq \beta} \|D^\alpha f\|_{L_2}^2} < \infty. \tag{1.10}$$

Sobolev spaces of non-integer smoothness $\beta \notin \mathbb{N}$ can be constructed in various equivalent ways ([10, 13, 49]). All spaces refer to a given domain $\mathcal{O}$, which may be dropped from the notation.

2. General method. In this section we present the skeleton of our approach, specialised to rates of contraction and uncertainty quantification. In the next sections, we verify the generic assumptions of the present section for combinations of priors and PDE models.

We assume given a prior distribution on the function $v=\mathcal{L}u_f$, and consider the two posterior distributions corresponding to the models (1.9) and (1.3):

(L) the posterior $\tilde\Pi_n(v\in\cdot\,|\,\tilde Y_n)$ of v in the observational model $\tilde Y_n = Kv + n^{-1/2}\,\dot{\mathbb{W}}$;
(N) the posterior $\Pi_n(f\in\cdot\,|\,Y_n)$ of f in the observational model $Y_n = u_f + n^{-1/2}\,\dot{\mathbb{W}}$.

The two models are linked through the identity $v=\mathcal{L}u_f$ and equations (1.5) and (1.8). Our approach is to construct the posterior distribution of v from model (L) and data $\tilde Y_n$, and next form the posterior distribution $\tilde\Pi_n\bigl(e(v)\in\cdot\,|\,\tilde Y_n\bigr)$ of $e(v)$ induced by the solution operator e given in (1.5). The latter equation shows that this is a posterior distribution of f in model (N) under the identification $\tilde Y_n = Y_n - \tilde g$.

To be more precise, a prior distribution on v that gives full probability to the set $\{\mathcal{L}u_f: f\in\mathcal{F}\}$, for some set $\mathcal{F}$, induces a prior on $f\in\mathcal{F}$ through (1.5). The induced posterior distribution $\tilde\Pi_n\bigl(e(v)\in\cdot\,|\,\tilde Y_n\bigr)$ of $e(v)$ is the ordinary posterior distribution $\Pi_n(f\in\cdot\,|\,Y_n)$ of f under this prior (see the proof of Proposition 2.1 for a precise statement and derivation of this correspondence). In particular, and trivially, this applies to priors on $v=\mathcal{L}u_f$ induced from a prior on f.

For flexibility we also allow prior distributions on v that are not restricted to the range of the map $f\mapsto\mathcal{L}u_f$, requiring only that the inverse map e:

(a) is defined on a set of full probability under the prior of v and
(b) recovers f as in (1.5) when applied to a function $\mathcal{L}u_f$.

The posterior distribution $\tilde\Pi_n\bigl(e(v)\in\cdot\,|\,\tilde Y_n\bigr)$ may then involve functions $f=e(v)$ for which existence of the forward solution u_f to the partial differential equation (1.4) may be hard to verify, or does not exist. In particular, PDE theory may ensure the existence of a solution u_f of (1.4) only under certain conditions on f.

For our purposes it is not important to verify such conditions; instead we can extend the definition of the forward map $f\mapsto u_f$ as follows, keeping the basic identities. For a given function v in model (L), the function $u=Kv+\tilde g$ is well defined and satisfies $v=\mathcal{L}u$ on $\mathcal{O}$ and $u=g$ on $\partial\mathcal{O}$, by (1.6)-(1.7). If $v=\mathcal{L}u_f$ for a function f and solution u_f of (1.4), then $u=u_f$ by (1.8) and $f=e(\mathcal{L}u_f)$ by the definition (1.5) of the solution operator. In the other case, if v does not belong to the range of the map $f\mapsto\mathcal{L}u_f$, then we *define* $f=e(v)$ and $u_f=u$ and again (1.5) and (1.8) hold, even if (1.4) may not. (It does if $c\bigl(e(v),Kv+\tilde g\bigr)=v$ on $\mathcal{O}$.) An alternative to extending the definition of u_f in this way is to consider the posterior distribution $\tilde\Pi_n\bigl(e(v)\in\cdot\,|\,\tilde Y_n\bigr)$ as a *projection-posterior* for f in the sense of [9], but this disregards that it is an ordinary posterior distribution in most cases.

The observations $\tilde Y_n$ and Y_n may be interpreted in terms of the white noise model, or be understood as vectors in $\mathbb{R}^n$, for given $x_{n,1},\ldots,x_{n,n}\in\mathcal{O}$, equal to the sum of the vectors $\bigl(Kv(x_{n,i}):i=1,\ldots,n\bigr)$ and $\bigl(u_f(x_{n,i}):i=1,\ldots,n\bigr)$ and a standard normal vector, in the linear and non-linear model, respectively.

2.1. *Posterior contraction rate.* We assume that the posterior distribution of v is a Borel law on a normed space $(V,\|\cdot\|)$ and possesses a rate of contraction ϵ_n to v_0 in the sense that $\tilde\Pi_n(v:\|v-v_0\|\geq\epsilon_n M_n|\,\tilde Y_n)\xrightarrow{P}0$, for any $M_n\to\infty$, under the assumption that v_0 gives the true distribution of $\tilde Y_n = Kv_0+n^{-1/2}\dot{\mathbb{W}}$.

PROPOSITION 2.1. *Suppose that the posterior distribution $\tilde\Pi_n(v\in\cdot\,|\,\tilde Y_n)$ of v in model (L) contracts under v_0 to $v_0=\mathcal{L}u_{f_0}$ at rate ϵ_n in $(V,\|\cdot\|)$ and satisfies $\Pi_n(v\in V_n|\,\tilde Y_n)\xrightarrow{P}1$ for given sets $V_n\subset V$. If (1.5) holds for a map e such that $e:V_n\to L_2$ is Lipschitz at v_0, then the posterior distribution of f in model (N) attains a rate of contraction ϵ_n under f_0 relative to the L_2-norm.*

PROOF. By assumption the prior distributions of $\mathcal{L}u_f$ and v are the same, and in view of (1.9) and (1.8), the conditional distributions of $Y_n - \tilde{g} = K\mathcal{L}u_f + n^{-1/2}\dot{\mathbb{W}}$ given $\mathcal{L}u_f$ and $\tilde{Y}_n = Kv + n^{-1/2}\dot{\mathbb{W}}$ given v are the same as well, under the identification $v = \mathcal{L}u_f$. It follows that the conditional distributions of $\mathcal{L}u_f$ given $Y_n - \tilde{g}$ and of v given $\tilde{Y}_n$ are the same. More precisely, if $(y, B) \mapsto L(y, B)$ is a Markov kernel such that $v|\,\tilde{Y}_n \sim L(\tilde{Y}_n, \cdot)$, then $\mathcal{L}u_f|\,Y_n \sim L(Y_n - \tilde{g}, \cdot)$.

By (1.5), we have that $f - f_0 = e(\mathcal{L}u_f) - e(\mathcal{L}u_{f_0})$ and hence $\Pi_n\big(f : \|f - f_0\|_{L_2} < \epsilon, \mathcal{L}u_f \in V_n|\,Y_n\big) = \Pi_n\big(f : \|e(\mathcal{L}u_f) - e(\mathcal{L}u_{f_0})\|_{L_2} < \epsilon, \mathcal{L}u_f \in V_n|\,Y_n\big)$. By the preceding paragraph this is the same as $L(Y_n - \tilde{g}, B)$, for $B = \{v \in V_n : \|e(v) - e(v_0)\|_{L_2} < \epsilon\}$. Since $Y_n - \tilde{g}$ is distributed as $\tilde{Y}_n$ under v_0, we obtain that $L(Y_n - \tilde{g}, B) \sim \tilde{\Pi}_n(B|\,\tilde{Y}_n)$. The assumptions that e is Lipschitz on V_n and that $\tilde{\Pi}_n(V_n|\,\tilde{Y}_n) \xrightarrow{P} 1$ show that there exists a constant C such that $\tilde{\Pi}_n(B|\,\tilde{Y}_n) \leq \tilde{\Pi}_n(v : \|v - v_0\| \leq C\epsilon|\,\tilde{Y}_n)$ on a set of probability tending to 1. The last probability tends to zero in probability for $\epsilon = M_n\epsilon_n$ and every $M_n \to \infty$. $\square$

The proposition uses the L_2-norm on the functions f for definiteness, and allows a general norm $\|\cdot\|$ on the functions $\mathcal{L}u_f$ for flexibility. It connects the posteriors for the two problems (1.4) and (1.9) in an abstract way, and needs to be made precise for particular problems. The input is a $\|\cdot\|$-contraction rate for v in the problem $\tilde{Y}_n = Kv + n^{-1/2}\,\dot{\mathbb{W}}$ at $v_0 = \mathcal{L}u_{f_0}$ and the guarantee that the posterior in this problem concentrates on sets V_n on which the solution map (1.5) is Lipschitz.

In our examples the Lipschitz assumption, relative to a natural norm, is usually mild. For instance, it can be verified by ascertaining that the posterior distribution $\tilde{\Pi}_n(Kv \in \cdot|\,\tilde{Y}_n)$ of Kv concentrates on functions that are bounded away from zero, or functions whose derivative is well behaved. We show this by establishing consistency of this posterior distribution for a relevant norm combined with the relevant assumption on the true function Kv_0. Because K is smoothing, uniform consistency for Kv is typically automatic for the usual priors, and often sufficient.

2.2. *Uncertainty quantification.* Credible sets for the linear inverse problem (L) map naturally into credible sets for the non-linear problem (N) of interest. Assume given a credible set $\tilde{C}_n(\tilde{Y}_n)$ for v based on the observation $\tilde{Y}_n = Kv + n^{-1/2}\dot{\mathbb{W}}$: a set of prescribed probability under the posterior distribution $\tilde{\Pi}_n(\cdot|\,\tilde{Y}_n)$. We transform this into the set of functions f given by

$$C_n(Y_n) := \big\{f : \mathcal{L}u_f \in \tilde{C}_n(Y_n - \tilde{g})\big\} = \big\{e(v) : v \in \tilde{C}_n(Y_n - \tilde{g})\big\}, \tag{2.1}$$

where e is the solution map given in (1.5). The set $C_n(Y_n)$ arguably has a non-standard shape. If balls are preferred, then one could consider instead its minimum enclosing ball, which has improved coverage and the same diameter.

The *credible level* of the credible set $\tilde{C}_n(\tilde{Y}_n)$ for v in the model $\tilde{Y}_n = Kv + n^{-1/2}\dot{\mathbb{W}}$ is by definition the posterior probability $\tilde{\Pi}_n\big(\tilde{C}_n(\tilde{Y}_n)|\,\tilde{Y}_n\big)$ and its *coverage* at v_0 is by definition the probability $\Pr_{v_0}\big(\tilde{C}_n(\tilde{Y}_n) \ni v_0\big)$ if $\tilde{Y}_n = Kv_0 + n^{-1/2}\dot{\mathbb{W}}$. For the model $Y_n = u_f + n^{-1/2}\dot{\mathbb{W}}$ the same quantities are defined analogously relative to the parameter f with true value f_0.

In the following proposition we identify $\tilde{Y}_n$ and $Y_n - \tilde{g}$, so that the assertions can be understood in an almost sure sense.

PROPOSITION 2.2. *The credible levels of the credible sets $\tilde{C}_n(\tilde{Y}_n)$ for v in model (L) and $C_n(Y_n)$ for f in model (N), given in (2.1), are equal, and so are the coverage levels of these sets at $v_0 = \mathcal{L}u_{f_0}$ and f_0, respectively. Furthermore, if the map $e : V_n \to L_2$ in (1.5) is uniformly Lipschitz at points $\bar{v}_n \in V_n$ and $\tilde{C}_n(\tilde{Y}_n) \subset V_n$, then on the event $\bar{v}_n \in \tilde{C}_n(\tilde{Y}_n)$ the L_2-diameters of the sets $C_n(Y_n)$ are of the same order in probability under f_0 as the $\|\cdot\|$-diameters of the sets $\tilde{C}_n(\tilde{Y}_n)$ under v_0.*

PROOF. As noted in the proof of Proposition 2.1, if the Markov kernel $(y,B)\mapsto L(y,\cdot)$ gives the posterior distribution of v given $\tilde Y_n=y$, then $L(Y_n-\tilde g,\cdot)$ gives the posterior distribution of $\mathcal{L}u_f$ given Y_n. By the definition of $C_n(Y_n)$ it follows that $\Pi_n\bigl(f\in C_n(Y_n)|\,Y_n\bigr)=\Pi_n\bigl(\mathcal{L}u_f\in\tilde C_n(Y_n-\tilde g)|\,Y_n\bigr)=L\bigl(Y_n-\tilde g,\tilde C_n(Y_n-\tilde g)\bigr)$, which is the same as $\tilde\Pi_n\bigl(\tilde C_n(\tilde Y_n)|\,\tilde Y_n\bigr)$, since $\tilde Y_n=Y_n-\tilde g$. Similarly $\Pr_{f_0}\bigl(f_0\in C_n(Y_n)\bigr)=\Pr_{f_0}\bigl(\mathcal{L}u_{f_0}\in\tilde C_n(Y_n-\tilde g)\bigr)$ is the same as $\Pr_{v_0}\bigl(v_0\in\tilde C_n(\tilde Y_n)\bigr)$.

The L_2-diameter $\sup\{\|f-g\|_{L_2}: f,g\in C_n(Y_n)\}$ of $C_n(Y_n)$ is equal to $\sup\{\|e(v)-e(w)\|_{L_2}: v,w\in\tilde C_n(\tilde Y_n)\}\le 2\sup\{\|e(v)-e(\bar v_n)\|_{L_2}: v\in\tilde C_n(\tilde Y_n)\}$, which is bounded by a constant times $\sup\{\|v-\bar v_n\|: v\in\tilde C_n(\tilde Y_n)\}$, because $\tilde C_n(\tilde Y_n)\subset V_n$ and e is uniformly Lipschitz at $\bar v_n\in V_n$. On the event $\bar v_n\in\tilde C_n(\tilde Y_n)$ the right side is bounded above by the diameter of $\tilde C_n(\tilde Y_n)$. □

Thus credible and coverage levels translate from the linear to the non-linear inverse problem for arbitrary credible sets. To retain also the size of the sets it may be necessary to restrict the starting sets $\tilde C_n(\tilde Y_n)$ to sets V_n on which the solution map (1.5) is locally Lipschitz.

In our examples the latter can be achieved without affecting credible or coverage levels. Indeed, in general as soon as $\mathbb{E}_{v_0}\tilde\Pi_n(V_n|\,\tilde Y_n)\to 1$ and $\Pr_{v_0}(V_n\ni v_0)\to 1$, the credible and coverage levels of the sets $\tilde C_n(\tilde Y_n)$ and $\tilde C_n(\tilde Y_n)\cap V_n$ are asymptotically the same. In our examples we choose the sets V_n for instance a uniform ball $V_n=\{v:\|Kv-K\hat v_n\|_\infty\le c_0\}$ of small radius around estimators $K\hat v_n$ of Kv_0. Then consistency of $K\hat v_n$ for Kv_0 for the uniform norm gives $\Pr_{v_0}(V_n\ni v_0)\to 1$, and consistency of the posterior distribution of Kv for Kv_0 relative to the uniform norm gives $\mathbb{E}_{v_0}\tilde\Pi_n(V_n|\,\tilde Y_n)\to 1$.

3. Bayesian analysis in the linear problem. In this section we review and extend results on the problem of recovering a function v from the noisy observation $\tilde Y_n=Kv+n^{-1/2}\dot{\mathbb{W}}$, where K is the operator given in (1.6). When applied to solving the non-linear problem (1.3), the function v will be taken equal to $v=\mathcal{L}u_f$. With a view towards this setting, we extend results from the literature by considering contraction relative to the uniform norm and smoothness norms. In Sections 3.1–3.3 we consider the white noise model, whereas in Section 3.4 we consider the discrete observational model. We restrict to mildly ill-posed problems in the sense of [8].

We consider Gaussian priors on v, or mixtures thereof, possibly with tuning parameters set by the empirical Bayes method. (Non-Gaussian priors were considered in [1, 33] and [18, 53].) Without loss of generality such a prior can be constructed by equipping the coefficients of v in a basis expansion with (conditionally) independent univariate normal distributions. For a given orthonormal basis (h_i) of $L_2(\mathcal{O})$ and $v=\sum_{i=1}^\infty v_ih_i$, we set

$$(v_1,v_2,\ldots)|\,\tau,\alpha\sim\bigotimes_{i=1}^\infty N(0,\tau^2 i^{-1-2\alpha/d}). \tag{3.1}$$

The hyper-parameters τ or α determine the smoothness of the prior, and may be chosen fixed, given a hyper prior, or set by the empirical Bayes method.

3.1. *Smoothness scales.* The basis (h_i) give rises to a scale of function classes G^s, for $s\ge 0$, consisting of all functions $v=\sum_{i=1}^\infty v_ih_i\in L_2(\mathcal{O})$ with $\|v\|_{G^s}<\infty$, for

$$\|v\|_{G^s}^2=\sum_{i=1}^\infty v_i^2 i^{2s/d}. \tag{3.2}$$

For $s<0$ we define the same norm on the sequences (v_i) and define G^s as the corresponding normed space; this can also be identified with the dual space of G^{-s}. Assume that the operator $K: G^0\to L_2(\mathcal{O})$ is *smoothing of order* p in the scale G^s in the sense that

$$\|Kv\|_{L_2}\asymp\|v\|_{G^{-p}},\qquad v\in G^0. \tag{3.3}$$

THEOREM 3.1. *Let $K : G^0 \to L_2$ be a linear operator satisfying (3.3). Consider the white noise model with the prior (3.1), for fixed τ and α.*

(i). *If $v_0 \in G^\beta$ and $-p \le \delta < \alpha\wedge\beta$, then the G^δ-contraction rate to v_0 of the posterior distribution of v is of the order $n^{-(\alpha\wedge\beta-\delta)/(2\alpha+2p+d)}$.*

(ii). *If either $\sup_i i^{u/d}\|Kh_i\|_\infty < \infty$, for some u with $d/2 < \alpha\wedge\beta+u$, or $K : G^{\gamma-p} \to G^\gamma$ is continuous and $\|\cdot\|_{H^\gamma(\mathcal{O})} \lesssim \|\cdot\|_{G^\gamma}$ for some γ with $d/2 < \gamma < \alpha\wedge\beta+p$, then the posterior distribution of Kv is consistent for Kv_0 relative to the uniform norm.*

PROOF. The contraction rate (i) is an extension of Theorem 6.1 in [18] to the $\|\cdot\|_{G^\delta}$-norm, proved in Theorem F.1.

For the uniform consistency (ii) of the posterior distribution of Kv under the first condition, we argue that $Kv = \sum_i v_i Kh_i$, and hence $\|Kv\|_\infty^2 \le \|v\|_{G^\delta}^2 \sum_i \|Kh_i\|_\infty^2 i^{-2\delta/d}$, by the Cauchy-Schwarz inequality. Under the condition on the uniform norms of the functions Kh_i, this is bounded above by a multiple of $\|v\|_{G^\delta}^2$, for $(2\delta+2u)/d > 1$, equivalently $\delta + u > d/2$. Thus the uniform consistency follows from the contraction in G^δ, which takes place for $\delta < \alpha\wedge\beta$. Under the condition $\alpha\wedge\beta+u > d/2$, a value of δ that satisfies both requirements exists. Under the second condition we use Sobolev embedding to see that $\|Kv\|_\infty \lesssim \|Kv\|_{H^\gamma(\mathcal{O})}$, for $\gamma > d/2$, which by assumption is bounded above by $\|Kv\|_{G^\gamma} \lesssim \|v\|_{G^{\gamma-p}}$. Thus the uniform consistency follows from contraction in G^δ for $\delta = \gamma - p$, which takes place if $\delta < \alpha\wedge\beta$. $\square$

The preceding theorem assumes fixed hyper-parameters. Theorem 7.2 of [18] shows that the mixture prior obtained by choosing a parameter τ from an inverse Gamma distribution and a fixed α gives an L_2-contraction rate $n^{-\beta/(d+2\beta+2p)}$, for any $\beta \in (0,\alpha]$, but has not been extended to rates in G^δ.

3.2. *Eigenbasis.* If $K : L_2(\mathcal{O}) \to L_2(\mathcal{O})$ is compact, then the self-adjoint operator $K^\intercal K : L_2(\mathcal{O}) \to L_2(\mathcal{O})$ possesses an orthonormal basis (h_i) of eigenfunctions. If κ_i^2 are the corresponding eigenvalues of $K^\intercal K$, then $\|Kv\|_{L_2}^2 = \langle K^\intercal Kv, v\rangle_{L_2} = \sum_{i=1}^\infty v_i^2\kappa_i^2$. Hence relative to the eigenbasis, K is smoothing of order p in the sense of (3.3) if

$$\kappa_i \asymp i^{-p/d}. \tag{3.4}$$

The sequence $(\psi_j)_{j\in\mathbb{N}}$ defined by $Kh_i = \kappa_i\psi_i$ forms an the orthonormal basis in the image space of K. If K is self-adjoint, then $K^s v = \kappa_i^s v$, and K is also smoothing in the extended sense that $\|Kv\|_{G^s} \asymp \|v\|_{G^{s-p}}$, for every $s \in \mathbb{R}$.

For the prior constructed relative to the eigenbasis, the white noise problem is equivalent to observing a sequence $(\tilde Y_{n,1}, \tilde Y_{n,2}, \ldots)$ of independent normal variables with $\tilde Y_{n,i}|\, v \sim N(\kappa_i v_i, 1/n)$, equipped with conditionally independent priors $v_i|\,\alpha,\tau \sim N(0, \tau^2 i^{-1-2\alpha/d})$. Conditionally on the hyper-parameters the problem is conjugate and the posterior can be obtained explicitly. The contraction rates for fixed hyper-parameters follow from Theorem 3.1, but contraction rates for priors with hyper-parameters set by the empirical and hierarchical Bayes methods were studied in [22, 45], and credible balls and adaptive credible balls in [23, 44]. (The parameter α in [22, 23, 44, 45] is denoted by α/d in the present paper.)

The hierarchical Bayes method can be implemented with fixed τ and a prior on α with density λ on $(0,\infty)$ chosen such that, for every $c_1 > 0$ there exist $c_2 \ge 0$, $c_3 \in \mathbb{R}$ and $c_4 > 1$, with $c_3 > 1$ if $c_2 = 0$, such that for $\alpha \ge c_1$,

$$c_4^{-1}\alpha^{-c_3}e^{-c_2\alpha} \le \lambda(\alpha) \le c_4\alpha^{-c_3}e^{-c_2\alpha}. \tag{3.5}$$

The empirical Bayes method chooses α equal to the maximum likelihood estimator, restricted to $[0, \log n]$, based on the Bayesian marginal distribution of the observation $\tilde{Y}_n$. This can be seen to be

$$\hat{\alpha}_n = \operatorname*{argmin}_{\alpha} \sum_{i=1}^{\infty} \Big[\log\Big(1 + \frac{n}{i^{1+2\alpha/d}\kappa_i^{-2}}\Big) - \frac{n^2}{i^{1+2\alpha/d}\kappa_i^{-2} + n}\tilde{Y}_{n,i}^2\Big]. \tag{3.6}$$

Alternatively, the parameter α may be fixed and the parameter τ chosen by empirical Bayes or equipped with a prior, see [45].

The papers [23, 44] studied credible balls of the type, for $\hat{v}_n$ the posterior mean and r_n set to the smallest value such that $\{v : \|v - \hat{v}_n\|_{L_2} \le r_n\}$ possesses a prescribed posterior probability in $(0, 1)$,

$$\tilde{C}_n(\tilde{Y}_n) = \{v : \|v - \hat{v}_n\|_{L_2} \le cr_n\}. \tag{3.7}$$

In [23] priors with fixed hyper-parameters (τ, α) were shown to give frequentist coverage provided α undershoots the true smoothness. As it is impossible to construct confidence sets of uniform coverage and rate-adaptive size, see e.g. [5, 37], neither the empirical nor the hierarchical Bayes methods can provide credible sets with prescribed uniform frequentist coverage over all possible parameters. However, in [38, 41, 44] the credible sets $\tilde{C}_n(\tilde{Y}_n)$ are shown to be confidence sets for true parameters whose coefficients decrease not too irregularly, so-called *polished tail* sequences.

We record these results, together with extensions of contraction in G^δ and uniform consistency, in the following theorem.

THEOREM 3.2. *Consider the white noise model with the prior based on the eigenbasis with coefficients* (3.1) *with fixed* τ *and* α *determined by the hierarchical or empirical Bayes method* (3.5) *or* (3.6).

(i). *If* $v_0 \in G^\beta$ *and* $-p \le \delta < \beta$, *then the* G^δ*-contraction rate of the posterior distribution of* v *is* $l_n n^{-(\beta-\delta)/(2\beta+2p+d)}$, *for* $l_n = (\log n)^{3/2}(\log\log n)^{1/2}$.
(ii). *If either* $\sup_i i^{u/d}\|Kh_i\|_\infty < \infty$, *for some* u *with* $d/2 < \beta + u$, *or* $\|\cdot\|_{H^\gamma(\mathcal{O})} \lesssim \|\cdot\|_{G^\gamma}$ *for some* γ *with* $d/2 < \gamma < \beta + p$, *then the posterior distribution of* Kv *is consistent for the uniform norm.*
(iii). *If the prior is constructed with fixed hyper-parameters* τ *and* α, *then the coverage tends to 1 provided* $\alpha < \beta$ *and the* L_2*-diameter is of the order* $O_P(n^{-(\alpha\wedge\beta)/(2\alpha+2p+d)})$.
(iv). *If* v_0 *satisfies the polished tail condition, then for sufficiently large* c *the coverage of the sets* (3.7) *tends to 1 and the* L_2*-diameter is of the order* $O_P(l_n n^{-\beta/(2\beta+2p+d)})$.

PROOF. Assertions (iii) and (iv) follow immediately from the results in [23] and [44]. Assertion (ii) follows from assertion (i) by the arguments in the proof of Theorem 3.1.

For the proof of (i), we extend the results of [22], which are based on the explicit representation of the posterior distribution: for $\lambda_i(\alpha) = i^{-1-2\alpha/d}$ and $\kappa_i^2 \asymp i^{-2p/d}$, the posterior distribution satisfies $v_i - v_{i,0} \,|\, (\tilde{Y}_n, \alpha) \sim N(m_i(\alpha), \sigma_i^2(\alpha))$, for

$$m_i(\alpha) := \frac{n\lambda_i(\alpha)\kappa_i\tilde{Y}_{n,i}}{1 + n\lambda_i(\alpha)\kappa_i^2} - v_{0,i}, \qquad \sigma_i^2(\alpha) := \frac{\lambda_i(\alpha)}{1 + n\lambda_i(\alpha)\kappa_i^2}.$$

Theorem 1 in [22] shows that the empirical Bayes estimator (3.6) is with probability tending to 1 under v_0 bounded below and above by deterministic sequences $\underline{\alpha}_n$ and $\overline{\alpha}_n$, characterised as crossing points of a certain deterministic criterion function h_n, derived from the likelihood. In the proof of Theorem 3 in the same reference, it is shown that in the hierarchical Bayes case the posterior probability that α falls in the interval $[\underline{\alpha}_n, \overline{\alpha}_n]$ tends in probability to one.

The lower bound satisfies $\underline{\alpha}_n \geq \beta - c_0/\log n$, for some constant c_0, which ensures that in both cases the rate of contraction does not fall essentially below the optimal rate for a β-smooth parameter.

It follows that, for the proof of (i) for both methods it suffices to show, for any $M_n \to \infty$,

$$\sup_{\underline{\alpha}_n \leq \alpha \leq \overline{\alpha}_n} \Pi_n\big(v : \|v - v_0\|_{G^\delta} > M_n l_n n^{-\frac{\beta-\delta}{d+2\beta+2p}} \,|\, \tilde{Y}_n, \alpha\big) \xrightarrow{P_{v_0}} 0. \tag{3.8}$$

By Markov's inequality, the posterior probability can be bounded above by $(M_n\epsilon_n)^{-2}\mathbb{E}\big(\|v - v_0\|^2_{G^\delta} \,|\, \tilde{Y}_n, \alpha\big)$, for ϵ_n the rate of contraction. The second moment can be computed using the explicit form of the posterior distribution. It follows that the preceding display is valid if the following two relations hold:

$$\sup_{\underline{\alpha}_n \leq \alpha \leq \overline{\alpha}_n} \sum_{i=1}^{\infty} m_i^2(\alpha) i^{2\delta/d} = O_{P_{v_0}}(l_n^2 n^{-\frac{2(\beta-\delta)}{d+2\beta+2p}}),$$

$$\sup_{\underline{\alpha}_n \leq \alpha \leq \overline{\alpha}_n} \sum_{i=1}^{\infty} \sigma_i^2(\alpha) i^{2\delta/d} = O(l_n^2 n^{-\frac{2(\beta-\delta)}{d+2\beta+2p}}).$$

The left side of the second equation is deterministic and monotonely decreasing in α, and hence it is sufficient to consider the series at $\alpha = \underline{\alpha}_n$. By elementary calculus, this can be seen to be of order $n^{-2(\underline{\alpha}_n-\delta)/(d+2\underline{\alpha}_n+2p)} \lesssim n^{-2(\beta-\delta)/(d+2\beta+2p)}$.

The analysis of the first relation is considerably harder, as it is random and non-monotone. However, the relation can be verified by the same arguments as in [22]. For completeness we include the complete proof in Section H of the supplement, see Theorems H.2 and H.3. □

In the preceding results the proof of uniform consistency is essentially based on Sobolev embedding. The following lemma shows that this approach is generally possible for parabolic or elliptic differential operators.

LEMMA 3.3. *Let (G^s) be the eigenscale of $K^T K$ for the inverse operator K (as in (1.6)) of a parabolic or elliptic differential operator $\mathcal{L}$ under the Dirichlet boundary condition. Then $K : G^\gamma \to L_\infty$ is continuous for every $\gamma > d$.*

PROOF. See Theorem B.2 in [21] for elliptic $\mathcal{L}$, and Theorem 4.5 in [51] (adapted to Dirichlet boundary condition) for parabolic $\mathcal{L}$. □

3.3. *Sobolev spaces.* The scale of Sobolev spaces $H^s(\mathcal{O})$ is of special interest in connection to differential operators. There are several bases that generate this scale, popular ones being the wavelet bases. Although a wavelet basis is most conveniently indexed by a double index in the form $(h_{j,k})$, the base functions can be ordered in a sequence and the resulting scale takes the form as in Section 3.1 (see Example F.1). The Gaussian prior on this sequence constructed as in Section 3.1 takes the form $\sum_j \sum_k v_{j,k} h_{j,k}$ in the wavelet domain, for independent variables $v_{j,k} \sim N(0, \tau^2 2^{-2j(\alpha+d/2)})$.

Thus we obtain the results of Theorem 3.1 for a wavelet prior, provided the operator K is smoothing (3.1) in the Sobolev scale. Unfortunately, this appears to be not often true for the full Sobolev scale, due to the boundary conditions connected to the differential operator. For instance, the inverse of the Laplacian is smoothing (of order 2) relative to the scale $H_0^1 \cap H^s(\mathcal{O})$, where H_0^1 is the space of weak solutions of (1.6), the subscript 0 arising from the Dirichlet boundary condition (see Proposition A.1). In such a case the contraction rates provided by Theorem 3.1 would still be valid for priors constructed from wavelet-like bases that observe the boundary conditions (or tight frames, see [50]). We conjecture that an unrestricted wavelet basis would give the same results, but a proof may be involved.

More abstractly, the operator K will be smoothing relative to the scale generated by the differential operator $\mathcal{L}$, and we may construct a Gaussian prior with covariance operator equal to the inverse of $\mathcal{L}$ (see Proposition 8.19 in [12]). The equivalence of Hilbert scales resulting from linear differential operators $\mathcal{L}$ and Sobolev spaces has been investigated extensively, see for instance Theorem 10.1 in [40] for elliptic differential operators $\mathcal{L}$.

3.4. *Discrete observations.* The preceding results for the white noise model extend to the discrete observational scheme, in which for given "design points" $x_{n,1},\dots,x_{n,n}\in\mathcal{O}$, we observe $\tilde{Y}_n(1),\dots,\tilde{Y}_n(n)$ given by, for i.i.d. standard normal variables Z_i,

$$\tilde{Y}_n(i) = Kv(x_{n,i}) + Z_i, \qquad i = 1,\dots,n.$$

The key is an interpolation technique, mapping discrete signals to the continuous domain, as employed in Chapter 8 of [53] and [54].

Denote the empirical inner product by $\langle v,w\rangle_{\mathbb{L}_n} := n^{-1}\sum_{i=1}^n v(x_{n,i})w(x_{n,i})$ and let $\|v\|_{\mathbb{L}_n} := \sqrt{\langle v,v\rangle_{\mathbb{L}_n}}$ be the induced semi-norm. Assume that for every $n\in\mathbb{N}$, there exists an n-dimensional subspace $\mathbb{L}_n\subset L_2(\mathcal{O})\cap C(\mathcal{O})$ such that, for constants $0<C_1<C_2<\infty$, independent of n,

$$C_1\|v\|_{L_2} \le \|v\|_{\mathbb{L}_n} \le C_2\|v\|_{L_2}, \qquad v\in\mathbb{L}_n. \tag{3.9}$$

Let $\mathcal{I}_n : L_2(\mathcal{O})\cap C(\mathcal{O})\to\mathbb{L}_n$ be the map such that $\mathcal{I}_n v$ is the unique element of $\mathbb{L}_n$ that interpolates v at the design points, i.e. $\mathcal{I}_n v(x_{n,i}) = v(x_{n,i})$, for every $i=1,\dots,n$. Then assume that for every s in some interval (s_d,S_d), and some sequence $\delta_{n,s}\to 0$,

$$\|\mathcal{I}_n v - v\|_{L_2} \lesssim \delta_{n,s}\|v\|_{G^s}. \tag{3.10}$$

THEOREM 3.4. *If* (3.9)-(3.10) *hold with* $n\delta_{n,\beta}^2\to 0$ *and* $s_d<\beta+p<S_d$ *and* $K: G^\beta\to G^{\beta+p}$ *continuous, then under the conditions of Theorem* 3.1 *the assertions (i)-(ii) of the theorem are valid also in the discrete observational model.*

PROOF. The assertion on the contraction rate is an extension of results of [53, 54] to smoothness norms, which is proved in Section G.

The remainder of the proof is the same as the proof of Theorem 3.1. □

4. Schrödinger equation. For a bounded domain $\mathcal{O}\subset\mathbb{R}^d$, a given non-negative function $f\in L_2(\mathcal{O})$ and a bounded, measurable function $g:\partial\mathcal{O}\to\mathbb{R}$, let u_f be the solution to the time-independent *Schrödinger equation*

$$\begin{cases} \frac{1}{2}\Delta u_f = f u_f, & \text{on } \mathcal{O},\\ u_f = g, & \text{on } \partial\mathcal{O}.\end{cases} \tag{4.1}$$

The goal is to recover the unknown "potential" f using a noisy observation of u_f.

This model can serve as a prototypical, benchmark example for elliptic PDEs, with substantial interest in its own right. Previous results in the literature placed a Gaussian prior distribution of fixed regularity on the (logarithm) of the functional parameter f. In [29] minimax convergence rates were derived for the MAP estimator corresponding to an optimally rescaled Gaussian process prior. Posterior contraction rates were derived for a uniform prior on the coefficients in a series expansion in [27] and for a rescaled Gaussian process prior in [25]. Frequentist coverage guarantees for credible balls in weak Sobolev spaces were deduced from uniform nonparametric Bernstein-von Mises results in [25, 27].

The solution u_f to the Schrödinger equation has a probabilistic expression through the Feynman-Kac formula (see Theorem 4.7 in [55]):

$$u_f(x) = \mathbb{E}^x \left[g(X_\tau) e^{-\int_0^\tau f(X_s)\, ds} \right], \qquad x \in \mathcal{O}. \tag{4.2}$$

Here $(X_s)_{s\geq 0}$ denotes a d-dimensional Brownian motion with starting value $X_0 = x \in \mathcal{O}$, and τ is its exit time from $\mathcal{O}$. It is known that $\sup_x \mathbb{E}^x \tau$ is finite (c.f. Proposition 1.16 in [55]). If f is bounded and g is bounded away from zero on $\partial\mathcal{O}$, then Jensen's inequality applied to (4.2) shows that

$$\inf_{x\in\mathcal{O}} u_f(x) \geq \inf_{y\in\partial\mathcal{O}} g(y) \inf_{z\in\mathcal{O}} e^{-\|f\|_\infty \mathbb{E}^z \tau} > 0.$$

This justifies an inversion formula of the form (1.5):

$$f = \frac{\Delta u_f}{2u_f}. \tag{4.3}$$

We conclude that to estimate f, it suffices to estimate Δu_f and u_f. In fact, an estimator for Δu_f suffices, since u_f can be recovered without error from Δu_f and the known boundary condition $u_f = g$ on $\partial\mathcal{O}$.

This motivates to take $\mathcal{L}$ as in (1.4) equal to the Laplacian $\mathcal{L} = \Delta$. Its inverse K as in (1.6) is then the inverse Laplacian under the Dirichlet boundary condition, and the function $\tilde{g}$ in (1.7) is the solution to the Laplace equation with boundary function g. We record the existence of this operator and function in the following lemma. We assume that the boundary points of $\mathcal{O}$ are *regular* in the sense of [16], page 25; a sufficient condition is that $\mathcal{O}$ satisfies the exterior sphere condition: for every $\xi \in \partial\mathcal{O}$ there exists a closed ball $B \subset \mathbb{R}^d$ such that $B \cap \bar{\mathcal{O}} = \{\xi\}$. This is true in particular for the domain $\mathcal{O} = (0,1)^d$.

LEMMA 4.1. *Let $\mathcal{L} = \Delta$ be the Laplacian. There exists a compact, self-adjoint linear operator $K : L_2(\mathcal{O}) \to L_2(\mathcal{O})$ such that (1.6) holds. If $\mathcal{O}$ is regular, and $g : \partial\mathcal{O} \to \mathbb{R}$ is continuous, then there exists a function $\tilde{g} : \mathcal{O} \to \mathbb{R}$ satisfying (1.7). If $\partial\mathcal{O}$ is C^{m+2}, for $m \in \mathbb{N} \cup \{0\}$, then $K : H^m(\mathcal{O}) \to H^{m+2}(\mathcal{O})$ is continuous.*

PROOF. Theorem 3 in Section 6.2 of [13] with L minus the Laplacian together with the example following the theorem show that there exists a unique (weak) solution $Ku \in H_0^1$ to equations (1.6) with $\mathcal{L}$ the Laplacian, for every $u \in L_2(\mathcal{O})$. The linearity of the map $u \mapsto Ku$ is clear, while compactness and self-adjointness is noted in the proof of Theorem 1 on page 335 of [13]. The continuity $K : H^m(\mathcal{O}) \to H^{m+2}(\mathcal{O})$ follows from Theorem 5 on page 323 of [13].

By Theorem 2.14 in [16], the Dirichlet problem on a bounded domain has a solution for every continuous boundary function g if and only if the boundary points of the domain are regular. By the remark made below the stated theorem, a sufficient condition for this is that the domain satisfies the exterior sphere condition. □

Thus we can proceed by the general approach outlined in the introduction with $\mathcal{L} = \Delta$ the Laplacian. Because $u_f = K\Delta u_f + \tilde{g}$, equation (4.3) shows that the solution map (1.5) can be written as $f = e(\Delta u_f)$ for

$$e(v) = \begin{cases} \frac{v}{2(Kv+\tilde{g})}, & \text{if } \operatorname{essinf} Kv + \tilde{g} > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{4.4}$$

LEMMA 4.2. *If $u_{f_0} = Kv_0 + \tilde{g}$ satisfies $\inf_{x\in\mathcal{O}} u_{f_0}(x) \geq c_0$, for some $c_0 > 0$, then the map $e : V \subset L_2(\mathcal{O}) \to L_2(\mathcal{O})$, for $V = \big\{v : \|Kv - Kv_0\|_\infty \leq c_0/2\big\}$, is Lipschitz at every uniformly bounded $\bar{v} \in V$ with Lipschitz constant bounded by a multiple of $1 \vee \|\bar{v}\|_\infty$.*

PROOF. If $Kv_0+\tilde g \geq c_0$ and $\|Kv-Kv_0\|_\infty \leq c_0/2$, then $Kv+\tilde g\geq c_0/2$. By the triangle inequality, uniformly in $v\in V$,

$$\big|e(v)-e(\bar v)\big| = \Big|\frac{v/2}{Kv+\tilde g} - \frac{\bar v/2}{K\bar v+\tilde g}\Big| \leq \frac{|v-\bar v|}{c_0} + \frac{|Kv-K\bar v|\,\|\bar v\|_\infty}{c_0^2}.$$

Thus the L_2-norm of left-side is bounded above by a multiple of $\|v-\bar v\|_{L_2} + \|Kv - K\bar v\|_{L_2}\|\bar v\|_\infty \lesssim \|v-\bar v\|_{L_2}(1\vee \|\bar v\|_\infty)$, by the continuity of K. □

THEOREM 4.3. *Suppose that $u_{f_0} = Kv_0+\tilde g$ is bounded away from zero and $\|v_0\|_\infty < \infty$. If the posterior distribution of Kv based on the observation $\tilde Y_n = Kv + n^{-1/2}\dot{\mathbb{W}}$ is consistent for the uniform norm, then the posterior distribution of f given Y_n contracts under f_0 at the same rate as the posterior distribution of v given $\tilde Y_n$ under v_0. Furthermore, the credible sets $C_n(Y_n)$ given in (2.1) have diameter of the same order in probability under f_0 as the credible sets $\tilde C_n(\tilde Y_n)$ under v_0 provided the latter sets are contained in $\{v : \|Kv - Kv_0\|_\infty < c_0/2\}$, for $c_0 = \inf_{x\in\mathcal{O}} u_{f_0}(x)$, and contain $\bar v_n$ with $\|\bar v_n\|_\infty = O_P(1)$.*

PROOF. We combine Lemma 4.2 with Proposition 2.1 to obtain the contraction rate, and with Proposition 2.2 to obtain the coverage of credible sets. □

Thus we focus on priors for v in the problem $\tilde Y_n = Kv + n^{-1/2}\dot{\mathbb{W}}$, where K is the inverse Laplacian under the Dirichlet boundary condition, that attain fast L_2-contraction rates and at the same time give uniformly consistent posteriors of Kv. Theorems 3.1, 3.2 and 3.4 give many possibilities. For a concrete example consider the domain $\mathcal{O} = (0,1)^d$.

It can be verified that the eigenfunctions of the inverse Laplacian on $\mathcal{O} = (0,1)^d$ with Dirichlet boundary condition are given by

$$h_{i_1,\ldots,i_j}(x_1,\ldots,x_d) = 2^{d/2}\prod_{j=1}^d \sin(i_j\pi x_j), \qquad (i_1,\ldots,i_d)\in\mathbb{N}^d, \tag{4.5}$$

with corresponding eigenvalues $-\kappa_{i_1,\ldots,i_d}$, for

$$\kappa_{i_1,\ldots,i_d} = \frac{1}{(\sum_{j=1}^d i_j^2)\pi^2}.$$

The indexing by the d-tuples $(i_1,\ldots,i_d)$ precludes an immediate application of contraction results for sequence priors. However the sequence $k_1\geq k_2\geq\cdots$ resulting from ordering the array of eigenvalues $(\kappa_i : i\in\mathbb{N}^d)$ in decreasing magnitude has polynomial decrease $k_\ell \asymp \ell^{-2/d}$, as $\ell\to\infty$ (see Lemma B.2 in Section B). Furthermore, if $v_1, v_2,\ldots$ are the array of coefficients $(\nu_i : i\in\mathbb{N}^d)$ in corresponding order of a function $v = \sum_{i\in\mathbb{N}^d}\nu_i h_i$, then the square norm (3.2) of the smoothness scale $(G^s)_{s\in\mathbb{R}}$ generated by the basis (h_i) ordered in a sequence satisfies, for $s\geq 0$,

$$\|v\|_{G^s}^2 = \sum_{l=1}^\infty v_l^2 l^{2s/d} \asymp \sum_{i\in\mathbb{N}^d}\nu_i^2\Big(\sum_{j=1}^d i_j^2\Big)^s.$$

Although the eigenfunctions are not always explicit, similar results are valid for the Laplacian on more general domains; see equation (3.1) in [19].

The next lemma verifies the interpolation assumptions of Theorem 3.4 (see Section B for a proof).

LEMMA 4.4. *Let (G^s) be the scale generated by the eigenfunctions (4.5) and let $\mathbb{L}_n := \{\sum_{i\in\mathbb{N}^d,\|i\|_\infty\le n^{1/d}} c_i h_i : c_i \in \mathbb{R}\}$. If $\beta > d/2$, then for every $v \in G^\beta$, there exists an element $\mathcal{I}_n v \in \mathbb{L}_n$ such that $\mathcal{I}_n v(x) = v(x)$ for every $x \in \{(\frac{2i}{2m+1})_{i=1,\dots,m}\}^d$, and*

$$\|\mathcal{I}_n v - v\|_{L_2} \lesssim n^{-\beta/d}\|v\|_{G^\beta}. \tag{4.6}$$

Furthermore, there exist constants $0 < C_1 < C_2 < \infty$ such that

$$C_1\|v\|_{L_2} \le \|v\|_{\mathbb{L}_n} \le C_2\|v\|_{L_2}, \qquad \forall v \in \mathbb{L}_n.$$

Given the eigenfunctions (h_i), for $i \in \mathbb{N}^d$, equip $v = \Delta u_f$ with the prior $\sum_{i=1}^\infty \nu_i h_i$, for $\nu_i \overset{\text{ind}}{\sim} N(0, \kappa_i^{d/2+\alpha})$, where the value of α is chosen either fixed or determined by the empirical Bayes or hierarchical Bayes approaches. For the array (h_i) ordered in by decreasing eigenvalues, this corresponds to a prior of the form $v = \sum_{\ell=1}^\infty v_\ell h_\ell$, with $v_\ell \overset{\text{ind}}{\sim} N(0, \ell^{-1-2\alpha/d})$, as in Section 3. To quantify the uncertainty of the procedure, consider credible sets of the form

$$C_n(Y_n) = \Big\{\frac{v}{2(Kv+\tilde{g})} : \|v - \hat{v}_n\|_{L_2} \le r_n, \|Kv - K\hat{v}_n\|_\infty \le \delta_0\Big\}, \tag{4.7}$$

for $\hat{v}_n$ the posterior mean of v, the constant r_n determined as the $1-\gamma$-quantile of the posterior distribution of $\|v - \hat{v}_n\|_{L_2}$ and a fixed constant $\delta_0 > 0$.

THEOREM 4.5. *Assume that $\Delta u_{f_0} \in G^\beta$, for $\beta > d/2$ and $\inf_x u_{f_0}(x) > 2\delta_0 > 0$. (i). The L_2-contraction rate of the posterior distribution of f to f_0 in the white noise model is $n^{-(\alpha\wedge\beta)/(d+2\alpha+4)}$ if the regularity of the prior is chosen deterministic and equal to α with $\alpha\wedge\beta+2 > d/2$. (ii). This contraction rate is $l_n n^{-\beta/(d+2\beta+4)}$ for a slowly varying sequence l_n, if α is chosen by the empirical Bayes or hierarchical Bayes methods and $\beta + 2 > d/2$. (iii). For the prior with deterministic α set to $\alpha = \beta - c/\log n$ and sufficiently large c, the credible set $C_n(Y_n)$ has frequentist coverage tending to one and L_2-diameter of the order $O_P(n^{-\beta/(d+2\beta+4)})$. (iv). For the prior with fixed $\alpha > d/2$, the contraction rate as in (i) is also attained in the discrete observational scheme.*

PROOF. The contraction rates (i) and (ii) in the Gaussian white noise model follow from combining Theorem 4.3 with Theorem 3.1 or Theorem 3.2(i)-(ii). Parts (i) of these theorems give a contraction rate for Δu_f, which translates into a contraction rate for f provided the uniform norms $\|u_f\|_\infty$ are bounded away from zero. The latter is guaranteed through the uniform posterior consistency obtained in parts (ii) of the two theorems, which apply with $u = 2$ under the conditions $\alpha\wedge\beta+2 > d/2$ and $\beta+2 > d/2$, respectively, as $\|Kh_\ell\|_\infty \lesssim \ell^{-2/d}$.

The frequentist coverage guarantee (iii) follows in the same way from combining Theorem 3.2(iii) with Proposition 2.2. The contraction rate (iv) in the discrete observational model follows from combining Lemma 4.4 and Theorem 3.4. □

As discussed in Section 3.3, similar results are valid for other priors than those resulting from the eigenexpansion, for instance priors based on certain wavelet expansions.

A sufficient condition for the assumption that the function Δu_{f_0} belongs to the Hilbert scale G^β is that it is compactly supported in $\mathcal{O}$ and belongs to the Sobolev space $H^\beta(\mathcal{O})$. See Proposition A.1 and Chapter 5, Appendix A (page 471), in particular equation (A.18) in [46].)

We show below that the contraction rate $n^{-\beta/(d+2\beta+4)}$ attained by the optimal, oracle scaling of the prior or via the empirical and hierarchical Bayes procedures are minimax rate optimal. We consider a more general setting, assuming that $c(f,u_f)=2fu_f$ (which is equivalent to assuming that the solution map is of the form (4.4)) and focus on the white noise model. This setting covers the Schrödinger model, and several others considered below. The proof of the theorem is deferred to Section B.1.

THEOREM 4.6. *Consider the Gaussian white noise model* (1.3) *under the PDE constraint* (1.4) *with* $\mathcal{L}=\Delta$ *and* $c(f,u_f)=2fu_f$. *Assume that the solution operator* K *of* (1.6) *is mildly ill-posed, with singular values satisfying* (3.4)*, and that the eigen basis* $(h_i)_{i\in\mathbb{N}}$ *of* K^TK *and the conjugate basis* $(Kh_i/\kappa_i)_{i\in\mathbb{N}}$ *are uniformly bounded. Then for fixed* $c_0, M>0$ *and* $\beta>d/2$,

$$\inf_{\hat f_n}\sup_{\mathcal{L}u_f\in G^\beta(M),\,u_f>c_0} E_f\|\hat f_n-f\|_2\gtrsim n^{-\beta/(d+2\beta+2p)}, \tag{4.8}$$

where G^β *denotes the eigenscale of* K^TK.

5. Heat equation with absorption. Recall the model 1.2. The discrete observational scheme version of this problem was investigated in [20], where minimax contraction rates and frequentist guarantees for credible sets in a weak metric space were derived under the assumption that the regularity β of the function $f\in H^\beta(\mathcal{O})$ is known.

The equation is of the form (1.4) for the parabolic differential operator $\mathcal{L}=\frac{\partial}{\partial t}-\frac12\Delta$, with $\Gamma=\big(\partial(0,1)^d\times[0,1]\big)\cup\big((0,1)^d\times\{0\}\big)$. The function f can be recovered from the solution u_f to the equation as

$$f=\frac{\mathcal{L}u_f}{u_f}.$$

The existence of the inverse operator $K\in L_2\big([0,1)^d\times[0,1]\big)$ satisfying (1.6) and the function $\tilde g$ satisfying (1.7) is proved in Theorems 3 and 4 of Section 7.1 in [13] and in Section 3 of [11]. This gives the solution map $f=e(\mathcal{L}u_f)$ as in (1.5) given by

$$e(v)=\begin{cases}\frac{v}{Kv+\tilde g}, & \text{if } \operatorname{essinf} Kv+\tilde g>0,\\ 0, & \text{otherwise.}\end{cases}$$

This is similar to the Schrödinger equation in Section 4, except for the different definition of K. Lemma 4 is true in exactly the same form for the present operator K and solution map.

In view of Proposition 2.1, an L_2-contraction rate of the posterior distribution of v to v_0 based on the observation $\tilde Y_n=Kv+n^{-1/2}\dot{\mathbb{W}}$ implies the same L_2-contraction rate for the posterior distribution of f, provided the posterior distribution of Kv is consistent for Kv_0 relative to the uniform norm and $u_{f_0}=Kv_0+\tilde g$ is bounded away from zero. Proposition 2.2 shows that the coverage and diameter of credible sets translate similarly. This is exactly as in Theorem 4.3, and applicable to many priors.

For a concrete example, consider the prior on the eigenfunctions of the operator $K^\intercal K$, which can be derived in closed form (see Lemma C.1). We may endow $v=\mathcal{L}u_f$ with a Gaussian prior $\sum_{i\in\mathbb{N}}v_ih_i$, for $h_1,h_2,\ldots$ the eigenfunctions ordered by decreasing eigenvalues and $v_i\overset{\text{ind}}{\sim}N(0,i^{-1-\alpha/d})$, with the hyper-parameter α either fixed to some value or selected with the empirical or hierarchical Bayes methods, as described in Section 3. For uncertainty quantification we consider the credible set described in (4.7).

Let (G^s) be the smoothness scale corresponding to the sequence of eigenfunctions, ordered by corresponding decreasing eigenvalues and viewed as functions on a $(d+1)$-dimensional domain. Thus a function belongs to G^β if it has a representation $v=\sum_{i\in\mathbb{N}}v_ih_i$ for coefficients $v_1,v_2,\ldots$ with $\sum_{i\in\mathbb{N}}i^{2\beta/(d+1)}v_i^2<\infty$.

THEOREM 5.1. *Assume that $\mathcal{L}u_{f_0} \in G^\beta$, for $\beta > d+1$, and $\inf_{(x,t)\in[0,1]^d\times[0,1]} u_{f_0}(x,t) > 2\delta_0 > 0$. (i). In the white noise model the L_2-contraction rate of the posterior distribution for f to f_0 is equal to $n^{-(\alpha\wedge\beta)/(d+1+2\alpha+2p)}$, with $p = 2(d+1)/(d+2)$, if the regularity of the prior is chosen deterministic and equal to $\alpha > d+1$, and equal to $l_n n^{-\beta/(d+1+2\beta+2p)}$ for a slowly varying sequence l_n if α is chosen by the hierarchical or empirical Bayes methods. (ii). For the prior with deterministic α set to $\alpha = \beta - c/\log n$ and sufficiently large c, the credible set (4.7) has frequentist coverage tending to one and L_2-diameter of the order $n^{-\beta/(d+1+2\beta+2p)}$.*

PROOF. In view of Lemma B.2 and Lemma C.1, the eigenvalues k_ℓ^2 of $K^\intercal K$ ordered by decreasing magnitude satisfy $k_\ell \asymp \ell^{-2/(d+2)}$. Thus the operator K is smoothing as in (3.3) relative to the eigenscale (G^s) on a $(d+1)$-dimensional domain of order p such that $2/(d+2) = p/(d+1)$, i.e. $p = 2(d+1)/(d+2)$.

Consequently, Theorem 3.2 gives L_2-contraction at the rates $n^{-(\alpha\wedge\beta)/(d+1+2\alpha+2p)}$, which is the rate given in (i), and $\|\cdot\|_{G^\gamma}$-contraction at some rate, for every $\gamma < \beta$.

By Proposition 2.1 the L_2-contraction rate of the posterior of v is transferred to the same rate for f provided that the posterior distribution of Kv is consistent for the uniform norm. By Lemma 3.3, the latter follows from consistency in the $\|\cdot\|_{G^\gamma}$-norm for some $\gamma > d+1$.

The frequentist coverage (ii) of the credible sets follows similarly, with the help of Proposition 2.2. □

REMARK 5.2. The space and time variables x and t in this problem do not play symmetric roles and it would be reasonable to treat them differently when estimating the function $(x,t) \mapsto f(x,t)$, in practice when devising a prior distribution and in theory when obtaining a contraction rate. In the preceding we lumped the two variables together in order to apply the contraction results on series priors given in Section 3. As shown in Lemma C.1, the singular values of $\mathcal{L}$ are of the order $\|i\|^2 + k$ (for the eigenfunctions $(x,t) \mapsto \prod_{i=1}^d \sin(i_j\pi x)\big(c_{i,k}\sin(\nu_{i,k}t) + \cos(\nu_{i,k}t)\big)$), where $\nu_{i,k} \asymp k$. This suggests that the problem is ill-posed of order 2 in the space variable and of order 1 in the time variable, as is also intuitively clear from the definition of $\mathcal{L}$. In the series setup of Section 3 this is accommodated by a single parameter of ill-posedness equal to $p = 2(d+1)/(d+2)$ (see the preceding proof), which is between 1 and 2. Using instead two parameters of ill-posedness would allow to obtain results for more general priors and more general model classes, more appropriate for the anisotropic nature of the problem. (A further refinement would be to model also the coordinates of the space variable anisotropically.) We leave this for further work.

REMARK 5.3. In [20] the function f was considered static over time, i.e. $f : \mathcal{O} \to \mathbb{R}$, while in our case f was allowed to change over time. We also note that our contraction rate is different than $n^{-\beta/(d+2\beta+4)}$ obtained in [20] for $f_0 \in H^\beta(\mathcal{O})$. However, both rates are minimax optimal in the respective function classes, i.e. for $\mathcal{L}u_{f_0} \in G^\beta$ and $f_0 \in H^\beta(\mathcal{O})$, respectively, see Theorem 4.6 with $p = 2(d+1)/(d+2)$ for the former one.

6. One-dimensional Darcy equation. For given functions $h : (0,1) \to \mathbb{R}$ and $g : \{0,1\} \to \mathbb{R}$, consider the one-dimensional differential equation

$$\begin{cases} f'u_f' + fu_f'' = h, & \text{on } (0,1), \\ u_f = g, & \text{on } \{0,1\}. \end{cases} \tag{6.1}$$

Because the solutions f to the first order linear ordinary differential equation $f'u' + fu'' = h$, for given u (and hence u' and u''), form a one-dimensional affine space, and we can at best retrieve u_f from the data, the function f can be recovered from the data at best up to a constant. We shall assume that the value $f(0)$ is given.

The equation can be written in the form $(fu_f')' = h$. Integration of this equation gives that $fu_f' - f(0)u_f'(0) = H$, for $H(x) = \int_0^x h(s)\,ds$. Under the assumption that u_f' is positive, we can solve f as $f = e(u_f')$ for

$$e(v)(x) = \frac{H(x) + f(0)v(0)}{v(x)}. \tag{6.2}$$

We conclude that we can write equation (6.1) in the form (1.4) with $\mathcal{L}u = u'$ (and $c(f, u_f) = (f(0)u_f'(0) + H)/f$) and solution map (1.5) given in the preceding display. Let $\tilde{g}$ be the function $\tilde{g}(x) = g(0) + \big(g(1) - g(0)\big)x$, and define an operator K by

$$Kv(x) = \int_0^x v(s)\,ds - x\int_0^1 v(s)\,ds.$$

Then $Kv(0) = Kv(1) = 0$, for every v, and hence $Ku' + \tilde{g} = g$ on $\partial\mathcal{O}$. Since also $(Ku' + \tilde{g})'' = u''$ on $\mathcal{O} = (0,1)$, it follows that $u_f = K\mathcal{L}u_f + \tilde{g}$, verifying the identity (1.8).

The adjoint of K is given by $K^\intercal u = -\int_0^{\cdot} u(s)\,ds + \int_0^1\int_0^t u(s)\,ds\,dt$, and hence $(K^\intercal Kv)'' = -(Kv)' = -v + \int_0^1 v(s)\,ds$, with (Neumann) boundary conditions $(K^\intercal Kv)'(x) = -Kv(x) = 0$, for $x \in \{0,1\}$. It follows that the eigenfunctions and corresponding eigenvalues of $K^\intercal K$ are given by

$$\begin{aligned} h_0 &= 1, \qquad \kappa_0^2 = 0,\\ h_i(x) &= \sqrt{2}\cos(\pi i x), \qquad \kappa_i^2 = \frac{1}{(\pi i)^2}, \qquad i \in \mathbb{N}. \end{aligned}$$

The boundary conditions $(u_f - \tilde{g})(0) = (u_f - \tilde{g})(1) = 0$ imply that $\int_0^1 (u_f - \tilde{g})'(s)\,ds = 0$, whence the functions $u_f' - \tilde{g}'$ are orthogonal to h_0. This motivates to consider the prior on $v = u_f'$ equal to the distribution of $\tilde{g}' + \sum_{i\in\mathbb{N}} v_i h_i$, for independent $v_i \sim N(0, i^{-1-2\alpha})$. Define $(G^s)_{s\in\mathbb{R}}$ as the scale generated by the eigenexpansion of $K^\intercal K$.

THEOREM 6.1. *Suppose $\|H\|_\infty < \infty$, $u_{f_0}' \in G^\beta$, for $\beta > 1/2$, and $\inf_{0<x<1} u_{f_0}'(x) > 0$. (i). For any $\alpha\wedge\beta > \delta > 1/2$ the contraction rate of the posterior distribution of f to $f(0)$ relative to the $\|\cdot\|_\infty$-norm is (at least) $n^{-(\alpha\wedge\beta-\delta)/(1+2\alpha+2)}$. (ii). If $f(0) = 0$ and $\alpha\wedge\beta > 1/2$, then the contraction rate of the posterior distribution of f to $f(0)$ relative to the L_2-norm is $n^{-(\alpha\wedge\beta)/(1+2\alpha+2)}$.*

PROOF. The solution map (1.5), relative to $\mathcal{L}u = u'$, is given by (6.2). On the set of functions $V_n = \{v : v \geq c\}$, for fixed $c > 0$, it satisfies

$$\big|e(v) - e(v_0)\big| \leq \frac{f(0)|v(0) - v_0(0)|}{c} + \big(\|H\|_\infty + f(0)v_0(0)\big)\frac{|v_0 - v|}{c^2} \lesssim \|v - v_0\|_\infty.$$

Since $v_0 = u_{f_0}'$ is bounded away from zero by assumption, the posterior distribution of v will concentrate on V_n as soon as it is consistent for the uniform norm. Thus in view of Proposition 2.1, for assertion (i) it suffices to prove that the posterior distribution of v contracts to v_0 relative to the uniform norm at the rate $n^{-(\alpha\wedge\beta-\delta)/(1+2\alpha+2)}$.

By Theorem 3.2 the posterior contraction rate for v relative to the $\|\cdot\|_{G^\delta}$-norm is $n^{-(\alpha\wedge\beta-\delta)/(1+2\alpha+2)}$. By construction $\langle u_{f_0}', 1\rangle_{L_2} = \langle v, 1\rangle_{L_2} = \tilde{g}(1) - \tilde{g}(0) = 0$, for almost every function v in a set of prior (and hence posterior) probability one. It follows that on this set

$$\|v - v_0\|_\infty = \sup_{0<x<1} \Big|\sum_{j\in\mathbb{N}} (v_j - v_{0,j}) h_j(x)\Big| \le \|v - v_0\|_{G^\delta} \sup_{0<x<1} \sqrt{\sum_{j\in\mathbb{N}} \frac{h_j(x)^2}{j^{2\delta}}}.$$

Since the functions h_j are uniformly bounded, the right side is bounded by a universal multiple of $\|v - v_0\|_{G^\delta}$, for any fixed $\delta > 1/2$. Thus the posterior contraction relative to the $\|v - v_0\|_{G^\delta}$-norm implies a posterior contraction relative to the uniform norm at the same rate.

To prove assertion (ii), we first note that if $f(0) = 0$, then the map e is Lipschitz at v_0 also relative to the L_2-norm on V_n. By Theorem 3.2 the posterior contraction rate for v relative to the L_2-norm is $n^{-(\alpha\wedge\beta)/(1+2\alpha+2)}$. By Proposition 2.1 this carries over to a posterior contraction rate for f provided that the posterior probability of the sets V_n tends to one. Under the assumption that $v_0 = u'_{f_0}$ is bounded away from zero, the latter follows from posterior consistency of v for the uniform norm. By the argument of the preceding paragraph this follows from consistency relative to the $\|\cdot\|_{G^\delta}$-norm, for some $\delta > 1/2$, and this is true for $\alpha\wedge\beta > 1/2$. □

REMARK 6.2. The assumption $\inf_{0<x<1} u'_{f_0}(x) > 0$ can be relaxed to the assumption that $\inf_{0<x<1}\big(|u'_{f_0}(x)| + u''_{f_0}(x)\big) > 0$. See Appendix D. The latter is the one-dimensional version of Assumption (7.3). See Section 7 for a brief discussion.

REMARK 6.3. The *two* boundary conditions of the non-homogeneous problem (6.1) cannot be removed by the general scheme (1.6) –(1.7) explained in the introduction when using the first order differential operator $\mathcal{L}u = u'$. Posing the problem instead in terms of the second order operator $\mathcal{L}_0 u = -u''$ would remedy this, and yield the same function $\tilde g$ and inverse operator $-KK^\intercal$ as in the preceding discussion. This would give $u_f = -KK^\intercal u''_f + \tilde g = Ku'_f + \tilde g$, and hence lead to the same inverse equation as introduced.

REMARK 6.4. Other boundary conditions than the Dirichlet boundary condition in (6.1) may motivate different priors. For instance, for the mixed boundary condition $u''_f(0) = u'_f(1) = 0$, we may use the same differential operator $\mathcal{L}u = u'$, but take the standard Volterra operator $Kv = \int_0^\cdot v(s)\,ds$ as its inverse. The eigenfunctions of $K^\intercal K$ are the functions $h_i(x) = \sqrt{2}\cos\big((i+1/2)\pi x\big)$ with corresponding eigenvalues $\kappa_i^2 = ((i+1/2)\pi)^{-2}$, for $i \in \mathbb{N}\cup\{0\}$. The functions in the eigenscale G^s generated by the operator $K^\intercal K$ satisfy the boundary condition on u'_f, and the preceding corollary remains valid provided $u'_{f_0} \in G^\beta$.

7. Darcy equation. Recall the model from (1.1). For a sufficiently smooth domain and sufficiently smooth functions h and g, a solution u_f is known to exist for every given positive $f \in H^\beta(\mathcal{O})$ with $\beta > 1 + d/2$ (see [28], Proposition 6.1.5 or Chapter 8 in [16]). Estimation of f in the discrete observational scheme version of this model was investigated in [17], who obtained contraction rates for certain Gaussian process priors on the function $\Phi^{-1}(f)$ for a given link function Φ.

For $d = 1$, this model reduces to the model considered in Section 6, which already revealed interesting features. The equation for $d > 1$ arises in many applied settings, but the challenges are substantial. The function f can be recovered from u_f only under certain conditions, and this may require that certain boundary values of f are pre-given, in addition to the boundary values on u_f in (1.1). Moreover, if recovery is possible, then the solution map (1.5) is not explicit, but only available as a numerical algorithm. Such difficulties can be avoided by taking multiple measurements of the solution function u_f, for the same function f, but with different boundary functions g. This is feasible in particular in an experimental setup, when the boundary function is under the control of the experimenter. Not only do multiple measurements simplify the inverse problem, but also better recovery rates may be expected. In this section we apply our general method with both a single and multiple measurements, in both cases with our operator $\mathcal{L}$ taken to be the Laplacian.

We start by a brief review of some aspects of the inversion map. A standard approach is the method of *characteristics* (see e.g. [13], p97–115), the set of curves $x:[\tau_0,\tau]\to\mathcal{O}$ in the domain $\mathcal{O}$ defined through the gradient flows $x'(t)=\nabla u\big(x(t)\big)$, with varying initial values $x(\tau_0)$. For each characteristic, equation (1.1) gives, for $u=u_f$,

$$\frac{d}{dt}f\big(x(t)\big)+f\big(x(t)\big)\Delta u\big(x(t)\big)=h\big(x(t)\big).$$

This ordinary differential equation can be solved by integrating factors, to give

$$f\big(x(t)\big)=e^{-\int_{\tau_0}^t \Delta u\left(x(s)\right)\,ds}f\big(x(\tau_0)\big)+\int_{\tau_0}^t h\big(x(s)\big)e^{-\int_s^t \Delta u\left(x(r)\right)\,dr}\,ds. \tag{7.1}$$

This equation completely specifies f on the characteristic $\{x(t):t\in[\tau_0,\tau]\}$ in terms of Δu and the starting value $f\big(x(\tau_0)\big)$. (It is attractive that the formula uses only the Laplacian Δu, which is the starting point of our reconstruction method, but the method also needs the gradient ∇u to construct the characteristics.) The method of characteristics is to cover the complete domain $\mathcal{O}$ by characteristics, and thus obtain an explicit solution to the inverse problem. This works best if the gradient ∇u never vanishes, which (given sufficient smoothness) would allow to extend every given characteristic (one can start one at every interior point of $\mathcal{O}$) to the boundary of $\mathcal{O}$. The domain is then covered by a set of characteristics starting at some boundary point (said to belong to the *influx boundary*) and exiting the domain at another boundary point. For a full reconstruction the function f then needs to be initialised on (only) the influx boundary.

A non-vanishing gradient is considered desirable, but depends on the problem, in particular on the boundary function g. If the gradient ∇u does vanish at some interior point of the domain, then a characteristic essentially stops there. It is shown in [35], that the method of characteristics is then still applicable under the condition that the Laplacian Δu is nonzero at such a point and has the same sign at all zeros of the gradient, i.e. $\Delta u+\|\nabla u\|>0$ throughout $\mathcal{O}$. In the latter case the domain $\mathcal{O}$ can be covered by characteristics as before, between boundary points, and characteristics starting at zeros of ∇u leading to the boundary (with infinite time set), and f can be recovered by integrating over the characteristics, as before. Interestingly, at a zero of ∇u, equation (1.1) specifies that $0+f\Delta u=h$, and hence the initial value of f at such a point is given by the value of $h/\Delta u$, which is measured and need not be externally specified. It may even be that all characteristics emanate from an interior point, or points, where the gradient vanishes (so that the influx boundary is empty) and no pre-given values of f are necessary for its recovery.

The influx boundary are the points where the gradient ∇u points into the interior of the domain $\mathcal{O}$. As seen previously, the recovery of f requires pre-given values on the influx boundary, and then f is determined on $\mathcal{O}$ and also at the remaining boundary points. This is somewhat tricky, as it indicates that one can also specify too many boundary values, and rule out every function f, the more so since the inflow boundary depends on u_f in general.

The paper [35] also discusses cases where there is no (unique) solution f to the inverse problem. In practice, such situations may be avoided if the measurements can be taken for specific boundary functions g. The inverse problem can be further simplified by taking measurements for multiple boundary functions. If $u_{f,1},\dots,u_{f,d}$ satisfy (1.1) for $h=0$

and boundary functions $g_1,\dots,g_d$, and the same function f, then (with $[a_1,\cdots a_d]$ the matrix with columns $a_1,\dots,a_d$),

$$\nabla f\cdot\big[\nabla u_{f,1}\cdots\nabla u_{f,d}\big]+f\big[\Delta u_{f,1}\cdots\Delta u_{f,d}\big]=0.$$

If the boundary functions can be chosen such that $\det[\nabla u_{f,1}\cdots\nabla u_{f,d}]\neq 0$ throughout the domain $\mathcal{O}$, then this gives the explicit inversion formula

$$\frac{\nabla f}{f}=-\big[\Delta u_{f,1}\cdots\Delta u_{f,d}\big]\big[\nabla u_{f,1}\cdots\nabla u_{f,d}\big]^{-1}. \tag{7.2}$$

For a convex domain $\mathcal{O}$, the function f (or rather $\log f$) can be recovered from this up to a multiplicative (or additive) constant by integration along lines starting from a single point. The construction of appropriate boundary functions g_i and other aspects of this approach are studied in [3].

7.1. *Single measurement.* In this section we consider the case of a single (noisy) measurement of the function u_f, following [28] and [17]. It is shown in [35] that every $u\in C^2(\bar{\mathcal{O}})$ such that

$$C(u):=\inf_{x\in\mathcal{O}}\big(\Delta u+\|\nabla u\|^2\big)(x)>0, \tag{7.3}$$

arises as $u=u_f$ for some function f that is absolutely continuous along the characteristics of u. [1] (The validity of Condition (7.3) can be monitored (at least asymptotically) within the context of our Bayesian procedure, as the posterior is supremum norm consistent, both for ∇u and Δu.) The following lemma, which is a slight adaptation of Proposition 2.1.5 of [28], shows that the inverse map $u_f\mapsto f$ is Lipschitz relative to the $H^2(\mathcal{O})$-norm, up to boundary values.

Let $\|f\|_{C^1(\mathcal{O})}=\|f\|_\infty+\|\nabla f\|_\infty$. A proof of the lemma, following [28], is included in Section E. (The assumed smoothness of the domain is needed only for validity of Green's formula; the lemma is also valid for e.g. $\mathcal{O}=[0,1]^2$.)

LEMMA 7.1. *Let $\mathcal{O}$ be a smooth bounded domain in $\mathbb{R}^d$ and let $g:\mathcal{O}\to\mathbb{R}$ and $h:\partial\mathcal{O}\to\mathbb{R}$ be given smooth functions. Let $u_f,u_{f_0}\in C^2(\mathcal{O})$ be solutions to (1.1) for given $f,f_0\in C^1(\mathcal{O})$. Then*

$$\begin{aligned}\|f-f_0\|^2_{L_2(\mathcal{O})}C(u_f)e^{-2\|u_f\|_\infty}\le 2\|f_0\|_{C^1(\mathcal{O})}\,\|u_f-u_{f_0}\|_{H^2(\mathcal{O})}\|f-f_0\|_{L_2(\mathcal{O})}\\+\int_{(\partial\mathcal{O})_{1,f}}(f-f_0)^2\,dS\sup_{x\in\partial\mathcal{O}}\|\nabla u_f(x)\|,\end{aligned}$$

where the integral in the last term is a surface integral over the inflow boundary $(\partial\mathcal{O})_{1,f}:=\{x\in\partial\mathcal{O}:\nabla u_f\cdot\vec{n}(x)<0\}$, for $\vec{n}$ the outer normal vector field on $\partial\mathcal{O}$. In particular, for every pair of positive functions $f,f_0\in H^\beta(\mathcal{O})$, for $\beta>1+d/2$, with $f=f_0$ on $\partial\mathcal{O}$,

$$\|f-f_0\|_{L_2(\mathcal{O})}\lesssim\frac{e^{2\|u_f\|_\infty}}{C(u_f)}\,\|\Delta u_f-\Delta u_{f_0}\|_{L^2(\mathcal{O})}. \tag{7.4}$$

[1] As noted in [35] and [17] condition (7.3) is certainly satisfied if f and h are strictly positive and bounded, since (1.1) implies $\inf h\le|\nabla\cdot(f\nabla u_f)|\le\|\nabla f\|\,\|\nabla u_f\|+f\Delta u_f$, whence either $2\|\nabla u_f\|\ge\inf h/\|\nabla f\|_\infty$ or $2\Delta u_f\ge\inf h/\|f\|_\infty$. In [35] it is also noted that weakening the condition by replacing Δu by its absolute value gives a very different problem, in which f is no longer uniquely determined.

Inequality (7.4) suggests that our approach with the operator $\mathcal{L} = \Delta$ equal to the Laplacian may yield contraction and coverage results relative to the $L_2(\mathcal{O})$-norm on the functions f. The multiplicative constant $e^{2\|u_f\|_\infty}/C(u_f)$ can be controlled by consistency of the posterior distribution of Δu_f relative to the uniform norm. Two potential difficulties are the boundary conditions on f and the fact that the formula is restricted to the surface of Laplacians of solutions u_f to (1.1).

Formula (7.4) presumes that f and f_0 agree on the influx boundary. The preceding discussion suggests that this presumption cannot be discarded and the first formula in the lemma shows that additional terms must be added otherwise. This is intrinsic to the inverse problem, but both practically and theoretically inconvenient.

A fortunate situation arises when the influx boundary is empty, in which case f can be reconstructed from u_f without specifying boundary values, and the inverse inequality (7.4) is valid. A case of interest where this is true, is when the boundary function g in (1.1) is constant and f is positive. Indeed, if $u_f = g$ is constant on $\partial\mathcal{O}$, and f is positive, then u_f cannot have a maximum in $\mathcal{O}$ by the maximum principle for elliptic partial differential equations ([16], page 31) and hence u_f will never increase when moving from a boundary point into the interior of $\mathcal{O}$, ensuring that $(\partial\mathcal{O})_{1,f}$ is empty.

Another case of interest is that the values of f on the full boundary happen to be known. Then one could proceed by constructing a posterior distribution that is concentrated on the set of functions u_f that agree with the known boundary values.[2] In [17] it is assumed that the function f is known and also constant on a neighbourhood of the boundary of $\mathcal{O}$. Under the same assumption, a posterior distribution on the Laplacian $v = \Delta u_f$, the starting point of our analysis, could be modified to correspond to functions f with the given boundary conditions before inverting to the function f by a solution operator (1.5). For instance, if the gradient ∇f vanishes near the boundary (as in [17]), then $\Delta u_f = h/f$ near the boundary, by (1.1), and a given function v can be modified into a function $\bar{v}$ that takes the known value h/f near the boundary. (This modification decreases the L_2-distance $\|v - \Delta u_{f_0}\|_{L_2(\mathcal{O})}$ as long as f_0 has the same boundary values.)

Another limitation of (7.4) is that it applies only to solutions u_f of (1.1). This can be circumvented, in principle, by defining a solution map $e : L_2(\mathcal{O}) \to L_2(\mathcal{O})$ by

$$e(v) = \operatorname*{argmin}_{f \in \mathcal{F}} \Big(f \mapsto \|\Delta u_f - v\|_{L_2(\mathcal{O})} : u_f \text{ solves } (1.1) \Big). \tag{7.5}$$

Boundary conditions on f can be inserted through the domain $\mathcal{F}$ of the minimisation. (We assume that the minimum is assumed, uniquely; otherwise, the argument can proceed with a consistent choice of a near minimiser, within a negligible tolerance.) If $v = \Delta u_f$ for $f \in \mathcal{F}$, then clearly $e(v) = f$, whence (1.5) is satisfied for $f \in \mathcal{F}$. Furthermore, by (7.4), provided $\exp(2\|u_{e(v)}\|_\infty)/C(u_{e(v)})$ is bounded,

$$\|e(v) - f_0\|_{L_2(\mathcal{O})} \lesssim \|\Delta u_{e(v)} - \Delta u_{f_0}\|_{L_2(\mathcal{O})} \le 2\|v - \Delta u_{f_0}\|_{L_2(\mathcal{O})},$$

in view of the triangle inequality and the fact that $\|v - \Delta u_{e(v)}\|_{L_2(\mathcal{O})} \le \|v - \Delta u_{f_0}\|_{L_2(\mathcal{O})}$, by the definition of $e(v)$. If $C(u_{f_0}) \ge c_0 > 0$ and $\|u_{f_0}\|_\infty \le d_0$, then consistency of the posterior distributions of Δu_f, ∇u_f and u_f for the uniform norm, shows that restricting $\mathcal{F}$ to functions such that $C(u_f) \ge c_0/2$ and $\|u_f\|_\infty \le 2d_0$ has an asymptotically negligible effect on the induced posterior distribution of $e(v)$. Then the L_2-contraction rate of a posterior distribution of the Laplacian $v = \Delta u_f$ is carried over into the same L_2-contraction rate of the induced posterior of the reconstructed function $f = e(v)$.

[2] Every positive $f \in H^\beta(\mathcal{O})$ defines a solution u_f. The solutions obtained from f with the given boundary values define a surface of functions that agree with these boundary values.

For a concrete example for the domain $\mathcal{O} = (0,1)^d$, we may use the prior on the eigen expansion of the Laplacian, as discussed in Section 4. Let (G^s) be the smoothness scale generated by the eigenfunctions (4.5). We assume that the boundary values of f are given and incorporated in such a way that (7.4) holds (see the preceding discussion). For instance, consider the case that g is constant with prior $v = \sum_{\ell=1}^\infty v_\ell h_\ell$, with $v_\ell \overset{\text{ind}}{\sim} N(0, \ell^{-1-2\alpha/d})$, or the case that f is equal to 1 in a neighbourhood of the boundary with the prior $v = \sum_{\ell=1}^\infty v_\ell h_\ell + h_0$, for a given function h_0 that takes the value h (for h as in (6.1)) on the boundary.

THEOREM 7.2. *Assume that $\Delta u_{f_0} \in G^\beta$ or $\Delta u_{f_0} \in G^\beta + h_0$, for $\beta > d/2$ and $C(u_{f_0}) > 0$. (i). The L_2-contraction rate of the posterior distribution of f to f_0 in the white noise model is $n^{-(\alpha\wedge\beta)/(2\alpha+4+d)}$ if the regularity of the prior is chosen deterministic and equal to α with $\alpha > d/2$. (ii). This L_2-posterior contraction rate is $l_n n^{-\beta/(2\beta+4+d)}$ for a slowly varying sequence l_n, if α is chosen by the empirical Bayes or hierarchical Bayes methods.*

PROOF. The contraction rates (i) and (ii) follow from Theorem 3.1 and Theorem 3.2, where we obtain both a contraction rate $n^{-(\alpha\wedge\beta-\delta)/(2\alpha+4+d)}$ for the Laplacian Δu_f relative to the G^δ norm and a contraction rate $n^{-(\alpha\wedge\beta)/(2\alpha+4+d)}$ for the Laplacian relative to the L_2-norm, with $\alpha = \beta$ in the case of (ii) and provided $\alpha \wedge \beta > \delta$. By Sobolev embedding, the first, with $\delta > d/2$, implies posterior consistency for the Laplacian relative to the uniform norm, thus allowing control over the constants $C(u_f)$ and $\|u_f\|_\infty$ needed to use the inversion formula 7.4 to obtain the contraction rate in L_2 for f from the contraction rate for the Laplacian. □

REMARK 7.3. The functions G^β in the closure of the eigenspace of the Laplacian vanish at the boundary and hence are contained in $H_0^\beta((0,1)^d)$ for large β. The conditions $\Delta u_{f_0} \in G^\beta$ and $C(u_{f_0}) > 0$ can then both be valid only if $\nabla u_{f_0} = \nabla K \Delta u_f + \nabla \tilde{g}$ does not vanish on the boundary. This appears to be typical, see p28 in [52], even though there do exist harmonic functions $\tilde{g}$ (as in (1.7) with $\mathcal{L} = \Delta$) such that $\nabla \tilde{g} = 0$ on a set of positive measure of the boundary ([4]). In the case that f is constant near the boundary, we use the prior offset h_0 to ensure that it can approximate Δu_{f_0} near the boundary. Independently, we note that positiveness of $C(u_f)$ on $\mathcal{O}$ may be relaxed to positiveness on a subset, such as all points in $\mathcal{O}$ bounded away its boundary, at the cost of replacing the $L_2(\mathcal{O})$-norm in the contraction rate to the L_2-norm restricted to the subset. It may be checked that the inversion estimate of Lemma 7.1 remains valid for this weakened L_2-norm under the relaxed condition.

While definition (7.5) proves the feasibility of our approach, it is not practically useful without an (efficient) algorithm to solve the minimisation problem. This is similar to solving the well known *deterministic* inverse problem of computing f from u_f, but better conditioned as our starting point is the Laplacian Δu_f instead of u_f. Given an estimate v for Δu_f, we can use (1.8) to compute $u = Kv + \tilde{g}$ and $\nabla u = \nabla K v + \nabla \tilde{g}$ as estimates of u_f and ∇u_f and next f by the method of characteristics (7.1). The first two steps are particularly easy if v is given in terms of the eigenexpansion $v = \sum_i v_i h_i$ of Δ, in which case $\nabla u = \sum_i \kappa_i v_i \nabla h_i + \nabla \tilde{g}$.

A numerical procedure implementing the third step (7.1) is already provided in [36], exactly under the condition that $C(u)$ is positive, but under the condition that the target function f is $C^2(\mathcal{O})$ (see Theorem 2 in [36]; note that f in the latter paper is our g and α is our f). In the remainder of this section we adapt the algorithm to take the Laplacian $v = \Delta u$ and gradient $w = \nabla u$ as inputs, and show that the inversion is stable if the function f is absolutely continuous (along discretised characteristics) and contained in $C^1(\mathcal{O})$. (We have not investigated whether the latter condition is always reasonable, in particular in the case of a vanishing gradient, but it cannot be much improved as the gradient ∇f appears in the defining equation. In Section D the assumption is verified in the case $d = 1$, even for a vanishing gradient, provided v is sufficiently smooth.)

We consider the case that the domain is the two-dimensional unit square $\mathcal{O} = [0,1]^2$. The construction can be extended to less regular domains along the lines described in Section 4 of [36]. For a given $\delta > 0$ so that $1/\delta \in \mathbb{N}$, consider the grid consisting of the points $x_{i,j} = (i\delta, j\delta)$, for $0 \le i, j \le 1/\delta$. It is shown in [35] that for given $u \in C^2([0,1]^2)$ satisfying (7.3), given values on the influx boundary $\partial\mathcal{O}_{1,u}$ and smooth functions g and h, there exists a function f that is absolutely continuous along the characteristics such that $u = u_f$, for u_f solving the Darcy equation (1.1) (and f is given by (7.1)). We shall give a numerically efficient algorithm to construct numbers $\alpha_{i,j}$ so that $\max_{i,j} |\alpha_{i,j} - f(x_{i,j})| \lesssim \delta^{\eta/2}$, for some $\eta > 0$, thus allowing the recovery of f at any precision. The complexity of the algorithm is linear in the number of grid points.

For a function $u : [0,1]^2 \to \mathbb{R}$, abbreviate $u(x_{i,j})$ to $u_{i,j}$. We are seeking an approximation to the solution f to the discretised equations $(\mathcal{L}_f u)_{i,j} = g_{i,j}$, for $\mathcal{L}_f$ the operator given by $\mathcal{L}_f u = \nabla f \cdot \nabla u + f\Delta u$. To this end, consider the approximation $\mathcal{L}_\alpha^\delta$, defined by

$$(\mathcal{L}_\alpha^\delta u)_{i,j} = \begin{cases} \alpha_{i,j}\Delta u_{i,j}, & \text{if } \|\nabla u_{i,j}\| < \sqrt{\delta}, \\ \frac{\|\nabla u_{i,j}\|}{\|z_{i,j}\|}(\alpha_{i,j} - \alpha_{k,l}) + \alpha_{i,j}\Delta u_{i,j}, & \text{if } \|\nabla u_{i,j}\| \ge \sqrt{\delta}, \end{cases} \tag{7.6}$$

where $(k,l) = (k_{i,j}, l_{i,j})$ are the coordinates of another grid point or a point in $\partial\mathcal{O}_{1,u}$ that is linked to $x_{i,j}$, as defined below, and $z_{i,j} = x_{i,j} - x_{k,l}$. The display mimics $(\mathcal{L}_f u)_{i,j} = \nabla f_{i,j} \cdot \nabla u_{i,j} + f_{i,j}\Delta u_{i,j}$, where f is replaced by α. In both cases in the display the term $f_{i,j}\Delta u_{i,j}$ is simply copied into $\alpha_{i,j}\Delta u_{i,j}$, but the term $\nabla f_{i,j} \cdot \nabla u_{i,j}$ is approximated. In the first case, when $\|\nabla u_{i,j}\| < \sqrt{\delta}$, the latter term is approximated by 0, while the second case involves a vector $z_{i,j}$, which is chosen in the direction of $\nabla u_{i,j}$, as follows. For $f \in C^1$ the approximation $f(x_{i,j}) - f(x_{i,j} - v) \approx \nabla f(x_{i,j}) \cdot v$, valid for any small vector $v \in \mathbb{R}^2$, gives $\nabla f_{i,j} \cdot \nabla u_{i,j} \approx \|\nabla u_{i,j}\|/\|z_{i,j}\|\big(f(x_{i,j}) - f(x_{i,j} - z_{i,j})\big)$, for a small vector $z_{i,j}$ in the direction of $\nabla u_{i,j}$. We arrive at the approximation in the display by choosing small $z_{i,j}$ in the direction of $\nabla u_{i,j}$ so that $x_{k,l} = x_{i,j} - z_{i,j}$ is a point of the grid or a point of the influx boundary. An explicit choice is

$$z_{i,j} = \delta\Big[\frac{1}{\sqrt{\delta}}\frac{\nabla u_{i,j}}{\|\nabla u_{i,j}\|}\Big], \tag{7.7}$$

where for a scalar x the notation $[x] = \operatorname{sign}(x)\lfloor |x| \rfloor$ denotes the integer closest to x inside the interval $[-|x|, |x|]$, and $[x] = ([x_1], [x_2])$ for a vector $x = (x_1, x_2)$. This choice always gives a point $x_{i,j} - z_{i,j}$ in the grid $\delta\mathbb{Z}^2$. If this point falls outside the unit square, then we redefine $z_{i,j}$ so that $x_{k,l} := x_{i,j} - z_{i,j}$ is the point in $\partial[0,1]^2$ where the line from $x_{i.j}$ to the original $x_{i,j} - z_{i,j}$ crosses this boundary. We shall refer to the latter type of points $x_{k,l}$ as belonging to the (discretised) influx boundary $\partial_d\mathcal{O}_{1,u}$ and assume that the values of f at these points are pre-specified.

For points $x_{i,j}$ in the influx boundary $\partial_d\mathcal{O}_{1,u}$ we set the value $\alpha_{i,j}$ equal to the pre-specified values $f(x_{i,j})$. For the other points we define $\alpha_{i,j}$ as the solution to the equation $(\mathcal{L}_\alpha^\delta u)_{i,j} = g_{i,j}$, thus mimicking the discretised Darcy equation $(\mathcal{L}_f u)_{i,j} = g_{i,j}$. In view of (7.6) this solution takes the form

$$\alpha_{ij} = \begin{cases} g_{i,j}/\Delta u_{i,j}, & \text{if } \|\nabla u_{i,j}\| < \sqrt{\delta}, \\ \frac{g_{i,j} + \alpha_{k,l}\|\nabla u_{i,j}\|/\|z_{i,j}\|}{\Delta u_{ij} + \|\nabla u_{i,j}\|/\|z_{i,j}\|}, & \text{if } \|\nabla u_{i,j}\| \ge \sqrt{\delta}. \end{cases} \tag{7.8}$$

Under the condition that $C(u)$ given in (7.3) is bounded away from zero, these quotients are well defined for small enough $\delta > 0$. In the first case of the display, the value $\alpha_{i,j}$ is expressed explicitly in the known values $g_{i,j}$ and $\Delta u_{i,j}$, but the second case involves the solution $\alpha_{k,l}$ at the grid point $x_{k,l}$, in addition to these known values and $\nabla u_{i,j}$. The latter recursion can be solved along a chain of recursions $x_{i',j'} \to x_{k',l'}$ that connects a point $x_{i,j}$ with $\|\nabla u_{i,j}\| \ge \sqrt{\delta}$ to a point x_{i_0,j_0} which either belongs to the influx boundary or satisfies $\|\nabla u_{i_0,j_0}\| < \sqrt{\delta}$. The value α_{i_0,j_0} is then determined and $\alpha_{i,j}$ can be computed by repeated substitutions. We show below that $u_{i,j} > u_{k,l}$ if $\|\nabla u_{i,j}\| \ge \sqrt{\delta}$, so that the equations are then solved in the order of increasing value of $u_{i,j}$. The strict decrease shows that the chain of recursions $x_{i',j'} \to x_{k',l'}$ cannot visit any grid point twice and hence must end at a point x_{i_0,j_0} after finitely many steps.

The proof of the following lemma is given in Section E.

LEMMA 7.4. *For $\mathcal{O} = [0,1]^2$ and given $u \in C^{2+\eta}(\mathcal{O})$ with $C(u) := \Delta u + \|\nabla u\|^2 > 0$ and $\eta \in (0,1]$ such that $u = u_f$ for $f \in C^1(\mathcal{O})$, the preceding algorithm gives $(\alpha_{i,j})$ such that, for small enough $\delta > 0$,*

$$\max_{0 \le i,j \le 1/\delta} |\alpha_{i,j} - f(x_{i,j})| \lesssim \delta^{\eta/2},$$

where the multiplicative constant depends on $\|u\|_{C^{2+\eta}(\mathcal{O})}$, $C(u)$, $\|f\|_{C^1(\mathcal{O})}$ and $\|g\|_{C^1(\mathcal{O})}$ only.

7.2. *Multiple measurements.* Suppose given multiple measurements $Y_{n,j} = u_{f,j} + n^{-1/2}\dot{\mathbb{W}}_j$, for $u_{f,1}, \dots, u_{f,d}$ solutions to (1.1) for $h = 0$ and boundary functions $g_1, \dots, g_d$. If the matrix $[\nabla u_{f,1} \cdots \nabla u_{f,d}]$ is invertible throughout the domain, then (7.2) gives an explicit reconstruction map for $\nabla f/f$. We can accommodate this in our setup by replacing the single measurement Y_n by the vector $(Y_{n,1}, \dots, Y_{n,d})$ and defining a solution operator e as in (1.5) as operating on the vector of Laplacians $\big(\Delta u_{f,1}, \dots, \Delta u_{f,d}\big)$, as follows.

If K is the inverse of the Laplacian $\mathcal{L} = \Delta$ as defined in (1.6), then $u_{f,j} = K\Delta u_{f,j} + \tilde{g}_j$ by (1.8), for $j = 1, \dots, d$, and hence $\nabla u_{f,j} = \nabla K \Delta u_{f,j} + \nabla \tilde{g}_j$. Our method is to infer $v_j = \Delta u_{f,j}$ from the inverse problem $Y_{n,j} - \tilde{g}_j = K v_j + n^{-1/2}\dot{\mathbb{W}}_j$. Then $\nabla K v_j + \nabla \tilde{g}_j$ is the corresponding estimator of $\nabla u_{f,j}$. This motivates to define the inverse operator

$$\bar{e}(v_1, \dots, v_d) = -\big[v_1, \dots, v_d\big]\big[\nabla K v_1 + \nabla \tilde{g}_1, \dots, \nabla K v_d + \nabla \tilde{g}_d\big]^{-1}.$$

By (7.2) and the preceding reasoning $\nabla f/f = e(\Delta u_{f,1}, \dots, \Delta u_{f,d})$. The map $\bar{e}$ is Lipschitz on the domain of functions $(v_1, \dots, v_d)$ such that the norm of the inverse matrices $\big[\nabla K v_1 + \nabla \tilde{g}_1, \dots, \nabla K v_d + \nabla \tilde{g}_d\big]^{-1}$ is uniformly bounded. We can compose $\bar{e}$ with a map that recovers f from $\nabla f/f$, given a boundary value, to construct a solution operator (1.5).

For a concrete example for the domain $\mathcal{O} = (0,1)^d$, we use the prior on the eigen expansion of the Laplacian, as discussed in Section 4. Let (G^s) be the smoothness scale generated by the eigenfunctions (4.5).

THEOREM 7.5. *Assume that $\Delta u_{f_0,1}, \dots, \Delta u_{f_0,d} \in G^\beta$, for $\beta > d/2 - 1$ and that the $(d \times d)$ matrix $\big[\nabla u_{f_0,1}(x) \cdots \nabla u_{f_0,d}(x)\big]$ is continuously invertible for every $x \in \mathcal{O}$ with inverses of uniformly bounded norms. (i). The L_2-contraction rate of the posterior distribution of $\nabla f/f$ to $\nabla f_0/f_0$ in the white noise model is $n^{-(\alpha\wedge\beta)/(2\alpha+4+d)}$ if the regularity of the prior is chosen deterministic and equal to α with $\alpha > d/2 - 1$. (ii). This L_2-posterior contraction rate is $l_n n^{-\beta/(2\beta+4+d)}$ for a slowly varying sequence l_n, if α is chosen by the empirical Bayes or hierarchical Bayes methods.*

PROOF. Theorems 3.1 and 3.2 give a posterior contraction rate $n^{-(\alpha\wedge\beta-\delta)/(2\alpha+4+d)}$ for the Laplacians $v_j = \Delta u_{f,j}$ relative to the G^δ norm and a posterior contraction rate

$n^{-(\alpha\wedge\beta)/(2\alpha+4+d)}$ for the Laplacians relative to the L_2-norm, with $\alpha=\beta$ in the case of (ii) and provided $\alpha\wedge\beta>\delta$.

From the eigen expansion of K (see (4.5)), we find that $\nabla K v=\sum_{i\in\mathbb{N}^d}\kappa_i v_i\nabla h_i$, where ∇h_i is the vector-valued function with kth coordinate $2^{d/2}\pi i_k\prod_{j\neq k}\sin(i_j\pi x_j)\cos(i_k\pi x_k)$ and $\kappa_i\asymp 1/\|i\|^2$. The latter functions without the factor πi_k are orthonormal in L_2 and hence $\|\nabla K v\|_{L_2(\mathcal{O})}\lesssim\|v\|_{L_2(\mathcal{O})}$ (easily). Furthermore, the same representation shows that $\|\nabla K v\|_\infty\le\sum_{i\in\mathbb{N}^d}\kappa_i|v_i|\,\|\nabla h_i\|_\infty\lesssim\sum_{i\in\mathbb{N}^d}\sqrt{\kappa_i}|v_i|$. Let $\tilde v_1,\tilde v_2,\ldots$ be the array of values $(v_i: i\in\mathbb{N}^d)$ ordered in a sequence by decreasing eigenvalues κ_i. Then in view of Lemma B.2, the last series can be bounded by a multiple of $\sum_{\ell=1}^\infty\ell^{-1/d}|\tilde v_\ell|\le(\sum_{\ell=1}^\infty\ell^{-(2+2\delta)/d})^{1/2}\|v\|_{G^\delta}$, by the Cauchy-Schwarz inequality. For $\delta>d/2-1$, the first series converges. We conclude that the operator ∇K is continuous both as an operator $\nabla K: L_2(\mathcal{O})\to L_2(\mathcal{O})$ and as an operator $\nabla K: G^\delta\to L_\infty(\mathcal{O})$.

The second shows that the posterior distribution of $\nabla K v_j+\nabla\tilde g_j$ is consistent for $\nabla u_{f_0,j}$ relative to the uniform norm. Since the matrices $\left[\nabla u_{f_0,1}\cdots\nabla u_{f_0,d}\right]$ are invertible uniformly throughout the domain $\mathcal{O}$, the same is true for the matrices $\left[\nabla K v_1+\nabla\tilde g_1\cdots\nabla K v_d+\nabla\tilde g_d\right]$, for every $v_1,\ldots,v_d$ in a set of posterior mass tending to one.

We can conclude that the map $\bar e: L_2(\mathcal{O})^d\to L_2(\mathcal{O})^d$ is Lipschitz on a set of posterior probability tending to one, so that the contraction rate of the posterior distributions of the Laplacians $v_j=\Delta u_{f,j}$ is carried over into the same contraction rate for the function $\nabla f/f$, by Proposition 2.1. ☐

REMARK 7.6. The function $\nabla f/f$ determines f up to a multiplicative constant, and the map $\nabla f/f\mapsto f$ is smooth provided f is bounded. Thus the contraction rates in the preceding theorem imply the same rates for f. One might hope that the rates for f are actually better, because of the integration involved in the map $\nabla f/f\mapsto f$. We have not investigated this further.

8. Acknowledgments. We thank Richard Nickl for sharing his insights during multiple discussions and pointing out references, in particular regarding the Darcy problem. The research leading to these results was partly financed by a Spinoza prize by the Netherlands Organisation for Scientific Research and co-funded by the European Union (ERC, BigBayesUQ, project number: 101041064). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

SUPPLEMENTARY MATERIAL

Supplementary Material to "Linear methods for non-linear inverse problems"
The Supplementary Material contains numerical studies for the proposed Bayesian linearization approach and an additional example covering the exponentiated Volter operator. Furthermore, the proofs of the theorems, lemmas, and corollaries from the main paper, along with additional technical lemmas, are also included.

REFERENCES

[1] AGAPIOU, S., LARSSON, S. and STUART, A. M. (2013). Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. *Stochastic Processes and their Applications* **123** 3828–3860.

[2] AGAPIOU, S. and SAVVA, A. (2024). Adaptive inference over Besov spaces in the white noise model using p-exponential priors. *Bernoulli* **30** 2275 – 2300.

[3] ALBERTI, G. S. and CAPDEBOSCQ, Y. (2018). *Lectures on elliptic methods for hybrid inverse problems* **25**. Société Mathématique de France, Paris.

[4] BOURGAIN, J. and WOLFF, T. (1990). A remark on gradients of harmonic functions in dimension ≥ 3. *Colloq. Math.* **60/61** 253–260. MR1096375

[5] CAI, T. T. and LOW, M. G. (2004). An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* **32** 1805–1840. MR2102494

[6] CASTILLO, I. (2012). A semiparametric Bernstein–von Mises theorem for Gaussian process priors. *Probab. Theory Related Fields* **152** 53–99. MR2875753

[7] CASTILLO, I. and NICKL, R. (2014). On the Bernstein-von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.* **42** 1941–1969. MR3262473

[8] CAVALIER, L. (2011). Inverse problems in statistics. In *Inverse problems and high-dimensional estimation*. *Lect. Notes Stat. Proc.* **203** 3–96. Springer, Heidelberg. MR2868199

[9] CHAKRABORTY, M. and GHOSAL, S. (2022). Rates and coverage for monotone densities using projection-posterior. *Bernoulli* **28** 1093–1119. MR4388931

[10] DI NEZZA, E., PALATUCCI, G. and VALDINOCI, E. (2012). Hitchhiker's guide to the fractional Sobolev spaces. *Bull. Sci. Math.* **136** 521–573. MR2944369

[11] DOOB, J. L. (1955). A probability approach to the heat equation. *Trans. Amer. Math. Soc.* **80** 216–280. MR79376

[12] ENGL, H. W., HANKE, M. and NEUBAUER, A. (1996). *Regularization of inverse problems*. *Mathematics and its Applications* **375**. Kluwer Academic Publishers Group, Dordrecht. MR1408680

[13] EVANS, L. C. (2010). *Partial differential equations*, second ed. *Graduate Studies in Mathematics* **19**. American Mathematical Society, Providence, RI. MR2597943

[14] FLEMMING, J. (2014). Regularization of autoconvolution and other ill-posed quadratic equations by decomposition. *J. Inverse Ill-Posed Probl.* **22** 551–567. MR3259109

[15] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. MR1790007

[16] GILBARG, D. and TRUDINGER, N. S. (2015). *Elliptic Partial Differential Equations of Second Order*. *Grundlehren der mathematischen Wissenschaften* **224**. Springer Berlin / Heidelberg, Berlin, Heidelberg.

[17] GIORDANO, M. and NICKL, R. (2020). Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. *Inverse Problems* **36** 085001, 35. MR4151406

[18] GUGUSHVILI, S., VAN DER VAART, A. and YAN, D. (2020). Bayesian linear inverse problems in regularity scales. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **56** 2081 – 2107.

[19] IVRII, V. (2016). 100 years of Weyl's law. *Bull. Math. Sci.* **6** 379–452.

[20] KEKKONEN, H. (2022). Consistency of Bayesian inference with Gaussian process priors for a parabolic inverse problem. *Inverse Problems* **38** 035002.

[21] KINDERLEHRER, D. and STAMPACCHIA, G. (2000). *An Introduction to Variational Inequalities and Their Applications*, first ed. *Classics in Applied Mathematics* **31**. Scoiety for Industrial and Applied Mathematics.

[22] KNAPIK, B., SZABO, B., VAN DER VAART, A. and VAN ZANTEN, H. (2016). Bayes procedures for adaptive inference in inverse problems for the white noise model. *Probability Theory and Related Fields* **164** 771-813.

[23] KNAPIK, B., VAN DER VAART, A. and VAN ZANTEN, H. (2011). Bayesian Inverse Problems with Gaussian Priors. *The Annals of Statistics* **39** 2626-2657.

[24] KNAPIK, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2013). Bayesian recovery of the initial condition for the heat equation. *Comm. Statist. Theory Methods* **42** 1294–1313. MR3031282

[25] MONARD, F., NICKL, R. and PATERNAIN, G. P. (2021). Statistical guarantees for Bayesian uncertainty quantification in nonlinear inverse problems with Gaussian process priors. *Ann. Statist.* **49** 3255–3298. MR4352530

[26] MONARD, F., NICKL, R. and PATERNAIN, G. P. (2021). Consistent inversion of noisy non-Abelian X-ray transforms. *Comm. Pure Appl. Math.* **74** 1045–1099. MR4230066

[27] NICKL, R. (2018). Bernstein-von Mises Theorems for statistical inverse problems I: Schrödinger Equation. *Journal of the European Mathematical Society* **22** 2697-2750.

[28] NICKL, R. (2023). *Bayesian Non-linear Inverse Problems*. *Zürich Lectures in Advanced Mathematics*. ETH, Zürich.

[29] NICKL, R., VAN DE GEER, S. and WANG, S. (2020). Convergence rates for penalized least squares estimators in PDE constrained regression problems. *SIAM/ASA Journal on Uncertainty Quantification* **8** 374–413.

[30] NICKL, R. and WANG, S. (2020). On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms.

[31] NIEMAN, D., SZABO, B. and VAN ZANTEN, H. (2023). Uncertainty quantification for sparse spectral variational approximations in Gaussian process regression. *Electronic Journal of Statistics* **17** 2250–2288.

[32] RANDRIANARISOA, T. and SZABO, B. (2023). Variational Gaussian processes for linear inverse problems. *Advances in Neural Information Processing Systems* **36** 28960–28972.

[33] RAY, K. (2013). Bayesian inverse problems with non-conjugate priors. *Electron. J. Statist.* **7** 2516–2549.

[34] RAY, K. (2017). Adaptive Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **45** 2511–2536. MR3737900

[35] RICHTER, G. R. (1981). An inverse problem for the steady state diffusion equation. *SIAM J. Appl. Math.* **41** 210–221. MR628945

[36] RICHTER, G. R. (1981). Numerical identification of a spatially varying diffusion coefficient. *Math. Comp.* **36** 375–386. MR606502

[37] ROBINS, J. and VAN DER VAART, A. (2006). Adaptive nonparametric confidence sets. *The Annals of Statistics* **34** 229–253.

[38] ROUSSEAU, J. and SZABO, B. (2020). Asymptotic frequentist coverage properties of Bayesian credible sets for sieve priors. *The Annals of Statistics* **48** 2155–2179.

[39] ROUSSEAU, J., SZABO, B. et al. (2017). Asymptotic behaviour of the empirical bayes posteriors associated to maximum marginal likelihood estimator. *The Annals of Statistics* **45** 833–865.

[40] SEELEY, R. T. (1965). Integro-Differential Operators on Vector Bundles. *Transactions of the American Mathematical Society* **117** 167-204.

[41] SNIEKERS, S. and VAN DER VAART, A. (2015). Adaptive Bayesian credible sets in regression with a Gaussian process prior. *Electron. J. Stat.* **9** 2475–2527. MR3425364

[42] SNIEKERS, S. and VAN DER VAART, A. (2020). Adaptive Bayesian credible bands in regression with a Gaussian process prior. *Sankhya A* **82** 386–425. MR4136240

[43] SZABO, B., HADJI, A. and VAN DER VAART, A. (2023). Adaptation using spatially distributed Gaussian Processes. *arXiv preprint arXiv:2312.14130*.

[44] SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428. MR3357861

[45] SZABÓ, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2013). Empirical Bayes scaling of Gaussian priors in the white noise model. *Electron. J. Stat.* **7** 991–1018. MR3044507

[46] TAYLOR, M. E. (2011). *Partial Differential Equations I: Basic Theory*. *Applied Mathematical Sciences*. Springer-Verlag, New York.

[47] TITSIAS, M. (2009). Variational learning of inducing variables in sparse Gaussian processes. In *Artificial intelligence and statistics* 567–574. PMLR.

[48] TRAVIS, L. and RAY, K. (2024). Pointwise uncertainty quantification for sparse variational Gaussian process regression with a Brownian motion prior. *Advances in Neural Information Processing Systems* **36**.

[49] TRIEBEL, H. (2008). *Function spaces and wavelets on domains*. *EMS Tracts in Mathematics* **7**. European Mathematical Society (EMS), Zürich. MR2455724

[50] TRIEBEL, H. (2008). *Function Spaces and Wavelets on Domains*, second ed. *EMS Tracts in Mathematics*. European Mathematical Society Publishing House.

[51] TRÖLTZSCH, F. (2010). *Optimal control of partial differential equations*. *Graduate Studies in Mathematics* **112**. American Mathematical Society, Providence, RI Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels. MR2583281

[52] WANG, S. (2026). Global polynomial-time estimation in statistical nonlinear inverse problems via generalized stability.

[53] YAN, D. (2020). *Bayesian Inference for Gaussian Models*. *PhD thesis*. University of Leiden, https://hdl.handle.net/1887/86070.

[54] YAN, D., GUGUSHVILI, S. and VAN DER VAART, A. (2024). Bayesian Linear Inverse Problems in Regularity Scales with Discrete Observations. *Sankhya A*.

[55] ZHAO, Z. and CHUNG, K. L. (2012). *From Brownian Motion to Schrodinger's Equation*. *Grundlehren der mathematischen Wissenschaften* **312**. Springer.