

Posterior consistency and convergence rates for Bayesian inversion with hypoelliptic operators

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2016 Inverse Problems 32 085005

(<http://iopscience.iop.org/0266-5611/32/8/085005>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

This content was downloaded by: ssiltanen

IP Address: 128.214.5.229

This content was downloaded on 24/07/2016 at 16:43

Please note that [terms and conditions apply](#).

Posterior consistency and convergence rates for Bayesian inversion with hypoelliptic operators

Hanne Kekkonen^{1,2}, Matti Lassas¹ and Samuli Siltanen¹

¹Department of Mathematics and Statistics PO Box 68, FI-00014 University of Helsinki, Finland

E-mail: hanne.kekkonen@helsinki.fi, matti.lassas@helsinki.fi and samuli.siltanen@helsinki.fi

Received 8 July 2015, revised 25 May 2016

Accepted for publication 2 June 2016

Published 23 June 2016



CrossMark

Abstract

The Bayesian approach to inverse problems is studied in the case where the forward map is a linear hypoelliptic pseudodifferential operator and measurement error is additive white Gaussian noise. The measurement model for an unknown Gaussian random variable $U(x, \omega)$ is $M_\delta(y, \omega) = A(U(x, \omega)) + \delta \mathcal{E}(y, \omega)$, where A is a finitely many orders smoothing linear hypoelliptic operator and $\delta > 0$ is the noise magnitude. The covariance operator C_U of U is smoothing of order $2r$, self-adjoint, injective and elliptic pseudodifferential operator. If \mathcal{E} was taking values in L^2 then in Gaussian case solving the conditional mean (and maximum *a posteriori*) estimate is linked to solving the minimisation problem $T_\delta(m_\delta) = \arg \min_{u \in H^r} \{ \|Au - m_\delta\|_{L^2}^2 + \delta^2 \|C_U^{-1/2} u\|_{L^2}^2 \}$. However, Gaussian white noise does not take values in L^2 but in H^{-s} where $s > 0$ is big enough. A modification of the above approach to solve the inverse problem is presented, covering the case of white Gaussian measurement noise. Furthermore, the convergence of the conditional mean estimate to the correct solution as $\delta \rightarrow 0$ is proven in appropriate function spaces using microlocal analysis. Also the frequentist posterior contractions rates are studied.

Keywords: Bayesian inverse problem, posterior consistency, white noise, convergence rate, microlocal analysis

(Some figures may appear in colour only in the online journal)

² Author to whom any correspondence should be addressed.

Notations

S^m	Class of pseudodifferential symbols of order m . See definition 1.
Ψ^m	Space of pseudodifferential operators (Ψ DO) of order m . See definition 2.
$H\Psi^{m,m_0}$	Space of hypoelliptic Ψ DOs of type (m, m_0) . See definition 3.
Ψ_p^m	Space of Ψ DOs of order m depending on spectral variable with order p . See definition 4.
$H\Psi_p^{m,m_0}$	Space of hypoelliptic Ψ DOs of type (m, m_0) depending on spectral variable with order p . See definition 5.
$Tr_{H^q \rightarrow H^q} C$	The trace of the operator $C : H^q \rightarrow H^q$. See (3.8).

1. Introduction

Practical inverse problems arise from the need to extract information from indirect data. For example, consider a device designed for measuring point values of a physical quantity $u(x)$. Technological imperfections cause the values of u at nearby points to merge together in the measurement. Mathematically this corresponds to convolution $\Phi * u$ by a point spread function Φ . The inverse problem is to recover the function u approximately from a finite number of point values of $\Phi * u$ corrupted by random white noise.

Computational inversion requires a finite representation of the quantity $u(x)$. In this paper we promote the view that it is a good idea to design a continuous model for u which then can be discretised with a desired number of degrees of freedom. This paves the way for the analysis of convergence as the discretisation becomes finer. Such convergence enables switching between different discretisations consistently; this is crucial for multigrid methods and for certain parameter-choice strategies.

In ill-posed inverse problems the measurement data alone is not sufficient for noise-robust recovery of the quantity of interest. For instance, Fourier transforming $\Phi * u$ gives $\widehat{\Phi} \widehat{u}$, so frequency-domain information is lost in areas where $\widehat{\Phi}$ is very close to zero. Therefore, successful computational inversion requires some *a priori* information in addition to the measurement data.

Practical inversion is all about combining measurement data and *a priori* information in a noise-robust way. The classical approach to do this is *regularisation* that assumes that the noise is deterministic and small in norm. Regularisation involves defining a family of continuous maps, parametrised by the norm $\|\text{noise}\|_{L^2}$, from the data space to the space of unknown quantities. This must be done so that as $\|\text{noise}\|_{L^2} \rightarrow 0$, the reconstruction approaches the true solution along a stable path. This methodology was originated by Tikhonov [53, 54]. Both continuous and discrete cases have been studied in depth in [9, 16, 24, 38, 40, 47, 55].

There is a serious drawback in the above noise model in the continuous limit. Namely, continuous white noise on \mathbb{R}^d is not square integrable. We discuss this in detail below in

section 3.1. The goal of this paper is to use *Bayesian inversion* to construct a consistent continuous-discrete framework covering the case of white noise.

Bayesian inversion is a flexible framework for combining measurement data and *a priori* information in the form of a *posterior distribution* [6, 10, 22, 33, 50]. Computational exploration of the finite-dimensional posterior distribution yields useful estimates of the quantity of interest and enables uncertainty quantification. Furthermore, analytic results about the continuous model can then be restricted to a given resolution in a discretisation-invariant way.

Our approach to Bayesian inversion follows a general strategy of computational mathematics: we consider a continuous model which can be discretised for any practical setting.

We study the following continuous model for indirect measurements:

$$M_\delta = AU + \mathcal{E}\delta, \quad (1.1)$$

where the random variables M_δ (data) and U (quantity of interest) take values in the Sobolev spaces $H^{-s}(N)$ and $H^\tau(N)$, respectively. Here N is a d -dimensional compact manifold e.g. a torus corresponding to a d -dimensional cube with opposite sides glued together. The real parameter τ is related to our *a priori* information about the smoothness of the unknown quantity of interest.

The measurement operator A in our model (1.1) is quite general: we assume it to be a finitely smoothing, injective hypoelliptic pseudodifferential operator (Ψ DO). See section 4 for precise definition. This class includes convolution operators with finitely smooth kernel Φ . One example of an operator that is hypoelliptic but not elliptic is the heat operator. For more examples of hypoelliptic operators see appendix A. The measurement noise \mathcal{E} is assumed to be normalised white Gaussian noise with mean zero and unit variance, and $\delta > 0$ models the noise amplitude.

We model practical measurement data by

$$M_k = P_k(AU) + P_k(\mathcal{E})\delta. \quad (1.2)$$

Here P_k is a linear operator related to measurement device; we assume that P_k is an orthogonal projection with k -dimensional range. We discretise the unknown U using some computationally feasible approximation of the form $U_n = S_n U$. Now we can study an inverse problem

$$\text{given a realisation of } M_k, \text{ estimate } U_n. \quad (1.3)$$

We are interested to know what happens to the approximated solutions of (1.1) when $\delta \rightarrow 0$. The analysis of small noise limit, also known as the theory of posterior consistency, has attracted a lot of interest in the last decade. Posterior convergence rates were first studied in [14, 48] and further in papers [1, 2, 7, 18, 21, 25, 26, 28, 42, 48, 51, 58]. However, much remains to be done. Developing a comprehensive theory is important since posterior consistency justifies the use of the Bayesian approach the same way as convergence results justify the use of regularisation techniques.

In the above mentioned papers the problem is studied from the frequentist point of view, that is, the data is thought to be generated by a fixed 'true' solution u^\dagger instead of random draw $U(\omega)$ from the prior distribution. This means that all the randomness in M comes from the randomness of the noise \mathcal{E} . The interest is then on the contraction of the posterior distribution around the 'true' solution u^\dagger as the noise goes to zero, see section 2.1.2 and theorem 3. The main emphasis of this paper, however, is in the purely Bayesian approach where it is assumed that also U is random. Since U and \mathcal{E} are assumed independent we can write

$$M_\delta(\omega) = AU(\omega_1) + \mathcal{E}(\omega_2)\delta, \quad (1.4)$$

where $\omega = (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$. In Bayesian case the posterior distribution is a function of (ω_1, ω_2) . Also the probability measure $d\mathbb{P}$ can be written in the following form

$$d\mathbb{P} = d\mathbb{P}_1(\omega_1)d\mathbb{P}_2(\omega_2).$$

We will denote the expected value over the joint distribution of U and \mathcal{E} by \mathbb{E} . The expected value over the noise is defined by

$$\mathbb{E}_U(F(\omega_1, \omega_2)) = \int_{\Omega_2} F(\omega_1, \omega_2)d\mathbb{P}_2(\omega_2). \quad (1.5)$$

Our paper provides a conceptual advantage over much of the existing literature. In many earlier studies A and C_U are perturbations of negative powers of operator $(I - \Delta)$. Our assumption, formulated in terms of hypoelliptic operators, means roughly speaking that the measurement operator A and the covariance operator C_U do not need to have a common basis in their singular value decomposition.

The rest of this paper is organised as follows. In section 2 we will introduce the Bayesian setting we are using and present our main result about convergence rates. In section 3 we take a closer look to the generalised Gaussian random variables in Sobolev spaces and introduce so called white noise paradox. We will also show that the distribution U takes values in H^τ , where τ is related to the smoothness of the solution and depends on the dimension of the space and the covariance of the prior. In section 4 we will introduce hypoelliptic operators and prove theorems 1 and 2. In section 5 we characterise credible sets and frequentist confidence regions and present and prove two theorems about the contraction of them. In appendix A we will give examples of some hypoelliptic operators and in appendix B we give a computational example.

2. Convergence results

Let us return to our indirect measurement problem

$$M_\delta = AU + \mathcal{E}\delta,$$

where we model $U = U(x, \omega)$, $M_\delta = M_\delta(y, \omega)$ and $\mathcal{E} = \mathcal{E}(y, \omega)$ as random functions. Here $\omega \in \Omega$ is an element of a complete probability space $(\Omega, \Sigma, \mathbb{P})$ and x and y denote the variables in domains of Euclidean spaces. The reason why we model also U as a random variable is that even though the unknown quantity is assumed to be deterministic we have only incomplete data of it. All information available about U before performing the measurements is included in a prior distribution that is independent of the measurement.

The Bayesian inversion theory is based on the Bayes formula. To solve the inverse problem (1.3) we have to express available *a priori* information of U_n in the form of a prior distribution π_{pr} in an n -dimensional subspace. Let M_k and \mathcal{E}_k be random vectors taking values in \mathbb{R}^k , and denote their distributions by π_{M_k} and $\pi_{\mathcal{E}_k}$, respectively. The solution of the inverse problem after performing the measurements is the posterior distribution of the unknown random variable. Given a realisation of the discrete measurement the posterior density for U_n taking values in the n -dimensional subspace is given by the Bayes formula

$$\begin{aligned}\pi(\mathbf{u} \mid \mathbf{m}_\delta) &= \frac{\pi_{\text{pr}}(\mathbf{u})\pi_{\mathcal{E}_k}(\mathbf{m}_\delta \mid \mathbf{u})}{\pi_{M_k}(\mathbf{m}_\delta)} \\ &= c\pi_{\text{pr}}(\mathbf{u})\exp\left(-\frac{1}{2\delta^2}\|\mathbf{m}_\delta - \mathbf{A}\mathbf{u}\|_2^2\right), \quad \mathbf{u} \in \mathbb{R}^n, \mathbf{m}_\delta \in \mathbb{R}^k,\end{aligned}\quad (2.1)$$

where $\mathbf{A} = P_kAS_n$ is a $k \times n$ matrix approximation to the operator A .

An approximated solution for the inverse problem is often given as a point estimate for (2.1). Let us assume that also U_n has Gaussian distribution. The maximum *a posteriori* (MAP) estimate $T_\delta^{\text{MAP}}: \mathbb{R}^k \rightarrow \mathbb{R}^n$ is defined by

$$T_\delta^{\text{MAP}}(M_k(\omega)) := \arg \max_{\mathbf{u} \in \mathbb{R}^n} \pi(\mathbf{u} \mid M_k(\omega)). \quad (2.2)$$

Note that the MAP estimate depends on ω through the realisation of the noise $\mathcal{E}_k(\omega)$ and unknown $U_n(\omega)$. When U_n is Gaussian distributed the MAP estimate coincides almost surely with the conditional (CM) mean estimate

$$T_\delta^{\text{CM}}(M_k(\omega)) = \mathbb{E}(U_n \mid \mathcal{M}_k)(\omega) \quad \text{a.s.}, \quad (2.3)$$

where \mathcal{M}_k is the σ -algebra generated by M_k .

Since in our case CM = MAP a.s. we will consider below the MAP estimate. Let us denote the covariance matrix of U_n by \mathbf{C}_{U_n} . Solving the maximisation problem (2.2) with a fixed realisation of noise and unknown corresponds to solving the minimisation problem

$$T_\delta(\mathbf{m}_\delta) = \arg \min_{\mathbf{u} \in \mathbb{R}^n} \left\{ \frac{1}{2\delta^2} \|\mathbf{A}\mathbf{u} - \mathbf{m}_\delta\|_2^2 + \frac{1}{2} \|\mathbf{C}_{U_n}^{-1/2}\mathbf{u}\|_2^2 \right\}. \quad (2.4)$$

Constructing S_n and π_{pr} is the core difficulty in Bayesian inversion. In many inverse problems there is no natural discretisation for the continuum quantity U , so n can be freely chosen. Consequently, S_n and π_{pr} should in principle be described for all $n > 0$. This raises the following questions: do the chosen S_n and π_{pr} represent the same *a priori* knowledge consistently at all resolutions $n > 0$? Does the estimate $T_\delta(\mathbf{m}_\delta)$ converge as $n \rightarrow \infty$? See e.g. [29, 34] Also, the number of data points may change, for example due to an updated measurement device. The aim of this paper is to build a rigorous theory that allows us to connect discrete models to their infinite-dimensional limit models in a consistent way.

We achieve consistent representation of *a priori* knowledge by constructing the prior distribution for U in the infinite-dimensional space X . Then the random variable $U_n = S_nU$ takes values in the finite-dimensional subspace $X_n \subset X$ and represents approximately the same *a priori* knowledge as U . The same way we construct distributions for M and \mathcal{E} in the infinite-dimensional space Y in which case the random variables M_k and \mathcal{E}_k take values in the finite-dimensional subspace $Y_k \subset Y$.

The finite-dimensional problem (2.4) Γ -converges as $n, k \rightarrow \infty$, under certain assumptions (including that m should be an L^2 -function), to the following infinite-dimensional minimisation problem in a Sobolev space H^r :

$$\arg \min_{u \in H^r} \left\{ \frac{1}{2\delta^2} \|m_\delta - Au\|_{L^2(N)}^2 + \frac{1}{2} \|C_U^{-1/2}u\|_{L^2(N)}^2 \right\}. \quad (2.5)$$

Above $C_U^{-1/2} \in \Psi^r$, that is, $C_U^{-1/2}$ is $-r$ orders smoothing pseudodifferential operator. See [23] for a proof. If we are thinking the above as a MAP estimate to a Bayesian problem we have to assume that U has formally the following distribution

$$\pi_{\text{pr}}(u) \underset{\text{formally}}{=} c \exp\left(-\frac{1}{2}\|C_U^{-1/2}u\|_{L^2(N)}^2\right).$$

Formula (2.5) only makes sense if the noise is square integrable. Even though

$$\|\varepsilon_k\|_{\ell^2} < \infty$$

with any $k \in \mathbb{N}$ the limit, when $k \rightarrow \infty$, is infinity. We will return to this so called ‘white noise paradox’ in section 3.1.

2.1. Main result

Let us now modify formula (2.5) to arrive at something useful for white Gaussian noise. When $\varepsilon \in L^2$ we can write

$$\|m_\delta - Au\|_{L^2(N)}^2 = \|Au\|_{L^2(N)}^2 - 2\langle m_\delta, Au \rangle_{L^2(N)} + \|m_\delta\|_{L^2(N)}^2. \quad (2.6)$$

Now omitting the infinite ‘constant term’ $\|m_\delta\|_{L^2(N)}^2$ we get a new minimisation problem which is well-defined also when m is not an L^2 function

$$T_\delta(m_\delta) := \arg \min_{u \in H^r(N)} \{ \|Au\|_{L^2(N)}^2 - 2\langle m_\delta, Au \rangle + \delta^2 \|C_U^{-1/2}u\|_L^2 \}, \quad (2.7)$$

where $\langle m_\delta, Au \rangle$ is interpreted as a suitable duality pairing instead of $L^2(N)$ inner product. When $A \in \Psi^{-t}$, $t \geq -\tau + s$, we can define $\langle m_\delta, Au \rangle = \langle m_\delta, Au \rangle_{H^{-s}(N) \times H^s(N)}$. Note that the forward operator A , the prior distribution and the noise depend on each other only through assumption $t \geq -\tau + s$.

It is well-known that the solution of the finite-dimensional problem (2.4) can be calculated using the following formula:

$$T_\delta(\mathbf{m}_\delta) = (A^T A + \delta^2 C_{U_n}^{-1})^{-1} A^T \mathbf{m}_\delta. \quad (2.8)$$

We can write the approximated solution $u_\delta := T_\delta(m_\delta)$ of the continuous problem (2.7) by

$$T_\delta(m_\delta) = (A^* A + \delta^2 C_U^{-1})^{-1} A^* m_\delta. \quad (2.9)$$

Before the main result of the paper we will study a simple example to give a reader an insight to Bayesian settings.

Example 1. Let N be a one-dimensional torus \mathbb{T}^1 . We are interested of the inverse problem

$$M_\delta = AU + \delta\mathcal{E},$$

where we assume that $\mathcal{E} \sim N(0, I)$ and $U \sim N(0, I)$, that is both the noise and the unknown are assumed to be normalised white Gaussian noise, see section 3 for rigorous definition. The white noise takes values in H^τ with some $\tau < -1/2$. On the other hand white noise has formally the following distribution

$$\pi_{\text{pr}}(u) \underset{\text{formally}}{=} c \exp\left(-\frac{1}{2}\|u\|_{L^2(\mathbb{T}^1)}^2\right).$$

Hence we want to solve

$$\arg \min_{u \in L^2} \{ \|Au\|_{L^2}^2 - 2\langle m_\delta, Au \rangle + \delta^2 \|u\|_{L^2(\mathbb{T}^1)}^2 \}.$$

Note that we are looking for a solution in $L^2(\mathbb{T}^1)$ even though the realisations of U are in $L^2(\mathbb{T}^1)$ with probability zero. In general if we are interested in finding a solution in $H^r(N)$ then we can show that the prior should take values in $H^\tau(N)$ where $\tau = r - s$, see section 3.

2.1.1. Convergence results in Bayesian setting. We will now formulate the main theorem of this paper about the convergence of the continuous solution (2.9). The precise definitions are discussed in more detail in section 3 and theorem 1 is proved in section 4.

Theorem 1. Let $r, s \in [0, \infty)$ and N be a d -dimensional closed manifold. Let $U(x, \omega)$ be a generalised Gaussian random function taking values in $H^\tau(N)$, $\tau = r - s$, with zero mean and covariance operator C_U . Assume that the operator C_U is a self-adjoint, injective and elliptic pseudodifferential operator (Ψ DO) of order $-2r$. Let $\mathcal{E}(y, \omega)$ be white Gaussian noise on N . Consider the measurement

$$M_\delta(y, \omega) = A(U(\cdot, \omega)) + \delta \mathcal{E}(y, \omega), \quad \omega \in \Omega,$$

where $A \in H\Psi^{-t, -t_0}$, $t > \max\{0, -\tau + s\}$ and $t \leq t_0 < 2t + r$, is a hypoelliptic pseudo-differential operator on the manifold N and $A : L^2(N) \rightarrow L^2(N)$ is injective. Above $\delta \in \mathbb{R}_+$ is the noise level and \mathcal{E} takes values in $H^{-s}(N)$ with some $s > d/2$.

Take $\zeta < \tau - 3(t_0 - t)$. Then we have the following convergence

$$\mathbb{E}\|U_\delta(\omega) - U(\omega)\|_{H^\zeta} \rightarrow 0, \quad \text{as } \delta \rightarrow 0. \quad (2.10)$$

The expectation is taken with respect to the joint distribution of (U, \mathcal{E}) . We have the following estimates for the speed of convergence:

(i) If $\zeta \leq t - s - 2t_0$ then there is such $C > 0$ independent of δ that

$$\mathbb{E}\|U_\delta(\omega) - U(\omega)\|_{H^\zeta} \leq C\delta^{\frac{2t-t_0+r}{t_0+r}}. \quad (2.11)$$

(ii) If $t - s - 2t_0 \leq \zeta < \tau - 3(t_0 - t)$ then there is such $C > 0$ independent of δ that

$$\mathbb{E}\|U_\delta(\omega) - U(\omega)\|_{H^\zeta} \leq C\delta^{-\frac{\zeta-\tau+3(t_0-t)}{t_0+r}}. \quad (2.12)$$

The different convergence speeds (i) and (ii) show the trade-off between the smoothness of the space and the speed of convergence. In case (i) we get better convergence rates but in case (ii) we can use a stronger norm. We see also that the smoother the forward operator A is the worse convergence rates we get.

We note that instead of the estimates (2.11) or (2.12), we could alternatively take the expected value only with respect to the noise in which case the constant C would depend on the realisation of $U(\omega)$. That is, proof of theorem 1 also shows that, we have almost surely

$$\limsup_{\delta \rightarrow 0} \frac{\mathbb{E}_U\|U_\delta(\omega) - U(\omega)\|_{H^\zeta}}{\delta^\nu} < \infty, \quad (2.13)$$

where $\nu = \frac{2t-t_0+r}{t_0+r}$ when $\zeta \leq t - s - 2t_0$ and $\nu = -\frac{\zeta-\tau+3(t_0-t)}{t_0+r}$ when $t - s - 2t_0 \leq \zeta < \tau - 3(t_0 - t)$.

Remark 1. The MAP-estimate U_δ takes values in the Cameron–Martin space of the prior. The Cameron–Martin space is the intersection of all linear subspaces where the random variable U belongs with probability one, and since there may be uncountable many such linear spaces, the Cameron–Martin space may be a zero measurable subset of the space where U takes values, see [3]. In the above settings where $U \sim N(0, C_U)$ and C_U is of order $-2r$ the random variable U takes values in H^τ , $\tau = r - s$ where $s > d/2$, and the Cameron–Martin space containing the MAP-estimate is H^r . In the Bayesian setting it is natural that the MAP-

estimate can not converge in a smaller space than the one U takes values. However, the same behaviour can be seen also in the deterministic setting when the unknown is in H^r and the MAP-estimate is thought as a Tikhonov regularised solution, see [23].

Example 2. Let us study a simple example in two-dimensional torus \mathbb{T}^2 . We consider a problem

$$M_\delta = (I - \Delta)^{-1}U + \delta\mathcal{E},$$

where \mathcal{E} is normalised white Gaussian noise that takes values in $H^{-s}(\mathbb{T}^2)$, $s > 1$ and $\delta > 0$ is the noise amplitude. The model operator $A = (I - \Delta)^{-1}$ is elliptic operator, smoothing of order 2. Let us consider the case when U has *a priori* distribution $N(0, C_U)$ where $C_U = (I - \Delta)^{-2}$, that is, $r = 2$. Then U takes values in $H^\tau(\mathbb{T}^2)$, where $\tau = r - s < 1$, almost surely and $U_\delta \in H^2$. Theorem 1 guarantees us convergence rate $C\delta$ when $\zeta \leq -3$ and (2.12) when $-3 \leq \zeta \leq 1$. For example we get the following convergence in $L^2(\mathbb{T}^2)$

$$\mathbb{E}\|U_\delta(\omega) - U(\omega)\|_{L^2(\mathbb{T}^2)} \leq C\delta^{\frac{1}{4}-\epsilon}$$

with $\epsilon > 0$ arbitrarily small.

2.1.2. Convergence results in frequentists setting. In the frequentist case one is often interested in the model

$$M_\delta^\dagger(\omega) = A(u^\dagger) + \delta\mathcal{E}(\omega), \quad (2.14)$$

where the data M_δ^\dagger is generated by a ‘true’ solution $u^\dagger \in H^\tau(N)$. Above \mathcal{E} is normalised white Gaussian noise and $\delta \in \mathbb{R}_+$ is the noise level. In (2.14) all the randomness of the M_δ^\dagger comes from the randomness of the noise \mathcal{E} . We denote

$$U_\delta^\dagger(\omega) = T_\delta(M_\delta^\dagger(\omega)). \quad (2.15)$$

In the frequentist setting we consider the case where u^\dagger is an arbitrary element of the space $H^\tau(N)$ where the random variable U takes values, instead of considering almost every element. Note that even though a set is measure-theoretically large, that is, has probability 1, it can still be topologically small (meager, or a set of first Baire category). For a discussion on these issues see [13].

We can then study the convergence $\mathbb{E}_{u^\dagger}\|U_\delta^\dagger(\omega) - u^\dagger\|_{H^\zeta}$ or in more frequentist spirit $\mathbb{E}_{u^\dagger}\|U_\delta^\dagger(\omega) - u^\dagger\|_{H^\zeta}^2$ where we use notation

$$\mathbb{E}_{u^\dagger}F(U_\delta^\dagger(\omega), u^\dagger, \omega) = \int_\Omega F(T_\delta(Au^\dagger + \delta\mathcal{E}(\omega)), u^\dagger, \omega)d\mathbb{P}(\omega), \quad (2.16)$$

that is, the expectation \mathbb{E}_{u^\dagger} is taken with respect to the noise $\mathcal{E}(\omega)$ and the other terms depending on ω , and u^\dagger is considered as a fixed parameter, c.f. (1.5). This means that after computing the estimator U_δ^\dagger using Bayesian methods we will consider the convergence of the estimator U_δ^\dagger to the ‘true’ solution u^\dagger which is not thought to be a random draw from the prior any more.

Remark 2. Next we consider the frequentist case when in addition it is assumed that $u^\dagger \in H^\tau(N)$ where $\tau \geq 0$. Note that then $r \geq \frac{d}{2}$. The mean integrated squared error (MISE) of an estimator \widehat{U} is defined

$$R(\widehat{U}, u^\dagger) = \mathbb{E}_{u^\dagger}\|\widehat{U} - u^\dagger\|_{L^2}^2. \quad (2.17)$$

The minimax risk $r_\delta(H^\tau(N))$ on the Sobolev space $H^\tau(N)$ is then given by

$$r_\delta(H^\tau(N)) = \inf_{\widehat{U}} \sup_{u^\dagger \in H^\tau(N)} R(\widehat{U}, u^\dagger),$$

where the infimum is taken over all estimators of the form $\widehat{U} = g(M_\delta^\dagger)$ where $g \in \mathcal{B}(H^{-s}, H^\tau)$. Here $\mathcal{B}(H^{-s}, H^\tau)$ is the set of Borel measurable functions from H^{-s} to H^τ .

Theorem 2. *Let $r > s > d/2$ and N be a d -dimensional closed manifold. Let $u^\dagger \in H^\tau(N)$ where $\tau = r - s > 0$. Assume that C_U , the covariance operator of the Gaussian prior, is a self-adjoint, injective and elliptic pseudodifferential operator of order $-2r$. Let $\mathcal{E}(y, \omega)$ be white Gaussian noise on N . Consider the measurement*

$$M_\delta^\dagger(y, \omega) = A(u^\dagger(\cdot)) + \delta \mathcal{E}(y, \omega), \quad \omega \in \Omega,$$

where $A \in H\Psi^{-t, -t_0}$, $t > \max\{0, -\tau + s\}$ and $t \leq t_0 \leq t + \tau/3$, is a hypoelliptic pseudodifferential operator on the manifold N and $A : L^2(N) \rightarrow L^2(N)$ is injective. Above $\delta \in \mathbb{R}_+$ is the noise level and \mathcal{E} takes values in $H^{-s}(N)$ with some $s > d/2$.

Then there is $C > 0$, independent of δ and u^\dagger , such that

$$\mathbb{E}_{u^\dagger} \|U_\delta^\dagger(\omega) - u^\dagger\|_{L^2}^2 \leq C(1 + \|u^\dagger\|_{H^\tau}^2) \delta^{\frac{2(\tau - 3(t_0 - t))}{t_0 + \tau}}. \tag{2.18}$$

Note that the assumptions on C_U in theorem 2 imply that C_U is a covariance operator of a random variable taking values in $H^\tau(N)$, see remark 3 below.

In the elliptic case $t = t_0$ we can write (2.18) as

$$\mathbb{E}_{u^\dagger} \|U_\delta^\dagger(\omega) - u^\dagger\|_{L^2}^2 \leq C(1 + \|u^\dagger\|_{H^\tau}^2) \delta^{\frac{2\tau}{s + \tau + \tau}}. \tag{2.19}$$

Since $s = \frac{d}{2} + \epsilon$ the above convergence rate (2.19) agrees, up to $\epsilon > 0$ arbitrarily small, with the minimax convergence rate, see [6]. The convergence of confidence regions is considered in section 5.

3. Generalised random variables

This section is largely based on the work of Lasanen [31, 32]; see also Piiroinen [41].

For any $s \in \mathbb{R}$, let $H^s(N)$ be the L^2 -based Sobolev space equipped with Hilbert space inner product

$$\langle \phi, \psi \rangle_{H^s(N)} = \int_N ((I - \Delta)^{s/2} \phi)(x) ((I - \Delta)^{s/2} \psi)(x) dx. \tag{3.1}$$

We also define a dual pairing between $H^{-s}(N)$ and $H^s(N)$

$$\langle \phi, \psi \rangle_{H^{-s}(N) \times H^s(N)} = \int_N \phi(x) \psi(x) dx \tag{3.2}$$

when $\phi, \psi \in C_0^\infty(N)$. Note that $H^0(N) = L^2(N)$. We often denote $H^s = H^s(N)$ and $L^2 = L^2(N)$.

A generalised Gaussian random variable V takes values in the space of generalised functions, and the pairing $\langle V, \phi \rangle$ with any test function $\phi \in \mathcal{D} = C^\infty(N)$ is a Gaussian random variable taking values in \mathbb{R}^d , see [45]. The generalised Gaussian random variables we will consider below are assumed to take values in some Hilbert space, typically in a Sobolev space $H^s(N)$, where the smoothness index $s \in \mathbb{R}$ may also be negative. Now, if V takes

values in H^s we say that V has the covariance operator $B_V : H^s \rightarrow H^s$ if

$$\mathbb{E}(\langle V - \mathbb{E} V, \phi \rangle_{H^s} \langle V - \mathbb{E} V, \psi \rangle_{H^s}) = \langle B_V \phi, \psi \rangle_{H^s}, \quad (3.3)$$

with any $\phi, \psi \in H^s$, see [45]. We can also define covariance operator to be a mapping $C_V : H^{-s} \rightarrow H^s$

$$\mathbb{E}(\langle V - \mathbb{E} V, \phi \rangle_{H^s \times H^{-s}} \langle V - \mathbb{E} V, \psi \rangle_{H^s \times H^{-s}}) = \langle C_V \phi, \psi \rangle_{H^s \times H^{-s}}, \quad (3.4)$$

with any $\phi, \psi \in H^{-s}$, see [3]. The connection between B_V and C_V is

$$B_V = C_V(I - \Delta)^s : H^s \rightarrow H^s.$$

Next we will take a closer look to the generalised white Gaussian noise and introduce the ‘white noise paradox’ by a simple example.

3.1. White noise paradox

White noise \mathcal{E} can be considered as a measurable map $\mathcal{E} : \Omega \rightarrow D'(N)$ where Ω is the probability space. Then normalised white noise is a random generalised function $\mathcal{E}(y, \omega)$ on N for which the pairings $\langle \mathcal{E}, \phi \rangle_{D' \times D}$ are Gaussian random variables for all test functions $\phi \in D = C^\infty(N)$, $\mathbb{E}\mathcal{E} = 0$, and

$$\mathbb{E}(\langle \mathcal{E}, \phi \rangle_{D'} \times D \langle \mathcal{E}, \psi \rangle_{D' \times D}) = \langle I\phi, \psi \rangle_{D' \times D} \quad \text{for } \phi, \psi \in D. \quad (3.5)$$

We will denote this by $\mathcal{E} \sim N(0, I)$. A realisation of \mathcal{E} is the generalised function $\varepsilon = \mathcal{E}(\cdot, \omega_0)$ on N with a fixed $\omega_0 \in \Omega$.

The probability density function of white noise \mathcal{E} is often *formally* written in the form

$$\pi_{\mathcal{E}}(\varepsilon) \stackrel{\text{formally}}{=} c \exp\left(-\frac{1}{2}\|\varepsilon\|_{L^2(N)}^2\right). \quad (3.6)$$

However, the realisations of the white Gaussian noise are almost surely not in $L^2(N)$. This brings us back to the problem in formula (2.5).

Example 3. Let \mathcal{E} be normalised white Gaussian noise defined on the d -dimensional flat torus $\mathbb{T}^d = (\mathbb{R}/(2\pi\mathbb{Z}))^d$. Let $e_{\vec{\ell}} \in L^2(\mathbb{T}^d)$, $\vec{\ell} = (\ell_1, \ell_2, \dots, \ell_d) \in \mathbb{Z}^d$ be an orthonormal basis of $L^2(\mathbb{T}^d)$ consisting of eigenfunctions of Laplacian, numbered so that $-\Delta e_{\vec{\ell}} = |\vec{\ell}|^2 e_{\vec{\ell}}$. Such functions $e_{\vec{\ell}}(x)$ can be chosen to be normalised products of the sine and cosine functions $\sin(\ell_j x_j)$ and $\cos(\ell_j x_j)$ that form the standard Fourier basis of $L^2(\mathbb{T}^d)$. The Fourier coefficients of \mathcal{E} with respect to this basis are independent, normally distributed \mathbb{R} -valued random variables with variance one, that is, $\langle \mathcal{E}, e_{\vec{\ell}} \rangle \sim N(0, 1)$. Then

$$\mathbb{E}\|\mathcal{E}\|_{L^2(\mathbb{T}^d)}^2 = \sum_{\vec{\ell} \in \mathbb{Z}^d} \mathbb{E}|\langle \mathcal{E}, e_{\vec{\ell}} \rangle|^2 = \sum_{\vec{\ell} \in \mathbb{Z}^d} 1 = \infty.$$

This implies that realisations of \mathcal{E} are in $L^2(\mathbb{T}^d)$ with probability zero. However, when $s > d/2$

$$\mathbb{E}\|\mathcal{E}\|_{H^{-s}(\mathbb{T}^d)}^2 = \sum_{\vec{\ell} \in \mathbb{Z}^d} (1 + |\vec{\ell}|^2)^{-s} \mathbb{E}|\langle \mathcal{E}, e_{\vec{\ell}} \rangle|^2 < \infty \quad (3.7)$$

and hence \mathcal{E} takes values in $H^{-s}(\mathbb{T}^d)$ almost surely (that is, with probability one).

On the other hand [44, theorem 2] implies that if $\|\mathcal{E}\|_{H^{-s}(\mathbb{T}^d)}^2 < \infty$ almost surely then $\mathbb{E}\|\mathcal{E}\|_{H^{-s}(\mathbb{T}^d)}^2 < \infty$ which yields $s > d/2$. This concludes that the realisations of white noise \mathcal{E} are almost surely in the space $H^{-s}(\mathbb{T}^d)$ if and only if $s > d/2$. In particular for $s \leq d/2$ the function $x \mapsto \mathcal{E}(x, \omega)$ is in $H^{-s}(\mathbb{T}^d)$ only when $\omega \in \Omega_0 \subset \Omega$ where $\mathbb{P}(\Omega_0) = 0$.

3.2. The smoothness of the prior

Consider the continuum measurement model $M_\delta = AU + \mathcal{E}\delta$ where the operator is now viewed as a smoothing map $A : H^\tau(N) \rightarrow H^{\tau+t}(N)$ for all $\tau \in \mathbb{R}$. We construct the prior by choosing U to be a generalised Gaussian random variable taking values in $H^\tau(N)$ and having expectation $\mathbb{E} U = 0$. First we will, however, give a definition for pseudodifferential operators.

Definition 1. Let $m \in \mathbb{R}$. We define the symbol class $S^m(\mathbb{R}^d, \mathbb{R}^d)$ to consist of $a(x, \xi) \in C^\infty(\mathbb{R}^d, \mathbb{R}^d)$ such that for all multi-indices α and β and any compact set $K \subset \mathbb{R}^d$ there is such constant $C_{\alpha, \beta, K} > 0$ that

$$|\partial_\xi^\alpha \partial_x^\beta a(x, \xi)| \leq C_{\alpha, \beta, K} (1 + |\xi|)^{m - |\alpha|}, \quad \xi \in \mathbb{R}^d, x \in K.$$

Definition 2. Let $Y : U \rightarrow \mathbb{R}^d$ be local coordinates of the manifold N . A bounded linear operator $A : \mathcal{D}'(N) \rightarrow \mathcal{D}'(N)$ is called a pseudodifferential operator if for any local coordinates $Y : U \rightarrow \mathbb{R}^d$, $U \subset N$, there is a symbol $a \in S^m(\mathbb{R}^d \times \mathbb{R}^d)$ such that for $u \in C_0^\infty(U)$ we have

$$Au(y_1) = \int_N k_A(y_1, y_2) u(y_2) dV_g(y_2),$$

where $k_A|_{N \times N \setminus \text{diag}(N)} \in C^\infty(N \times N \setminus \text{diag}(N))$ and $\text{diag}(N) = \{(y, y) \in N \times N \mid y \in N\}$. Also when $Y : U \rightarrow V \subset \mathbb{R}^d$ are local C^∞ -smooth coordinates $k_A(y_1, y_2)$ is given on $U \times U$ by

$$k_A(Y^{-1}(x_1), Y^{-1}(x_2)) = \int_{\mathbb{R}^d} e^{i(x_1 - x_2) \cdot \xi} a(x_1, \xi) d\xi,$$

where $x_1, x_2 \in V \subset \mathbb{R}^d$ and $a = a_V \in S^m(V, \mathbb{R}^d)$. In this case we will write

$$A \in \Psi^m(N),$$

and say that in local coordinates $Y : U \rightarrow V \subset \mathbb{R}^d$ the operator A has the symbol $a(x, \xi) \in S^m(V \times \mathbb{R}^d)$.

We assume that the covariance operator $C_U \in \Psi^{-2r}$, that is C_U is smoothing of order $2r$, self-adjoint and elliptic. With given $r \in \mathbb{R}$ we have to choose $\tau \in \mathbb{R}$ so that $B_U = C_U(I - \Delta)^\tau \in \Psi^{-2(r-\tau)}$ is a trace class operator. An operator B_U is in the trace class $\mathfrak{S}^1(H^\tau)$ if

$$\text{Tr}_{H^\tau \rightarrow H^\tau}(B_U) = \sum_{j=1}^{\infty} |\lambda_j| < \infty, \quad (3.8)$$

where λ_j are the eigenvalues of the operator B_U . Condition $B_U \in \mathfrak{S}^1(H^\tau)$ guarantees that $\mathbb{E}(U, U)_{H^\tau} < \infty$.

Let ν_j be the eigenvalues of $B_U^{-1} \in \Psi^{2(r-\tau)}$. Counting the geometric multiplicity of the eigenvalues, we arrange the eigenvalues of B_U^{-1} in ascending order as

$$\nu_1 \leq \nu_2 \leq \dots \leq \nu_k \leq \dots$$

Since B_U^{-1} is a self-adjoint elliptic operator with smooth coefficients Weyl's law for elliptic operators tells us that the number $N(\nu) = \#\{\nu_j \mid \nu_j \leq \nu\}$ of the eigenvalues of B_U^{-1} in a closed manifold less than or equal to ν has asymptotics

$$N(\nu) \sim c\nu^{\frac{d}{2(r-\tau)}}(1 + O(\nu^{-\frac{1}{2(r-\tau)}})) \quad \text{when } \nu \rightarrow \infty.$$

Hence for the eigenvalues λ of the operator $B_U \in \Psi^{-2(r-\tau)}$

$$\lambda_j \sim c'j^{-\frac{2(r-\tau)}{d}}(1 + o(1)) \quad \text{when } j \rightarrow \infty.$$

To satisfy condition (3.8) we require

$$\sum_{j=1}^{\infty} |\lambda_j| \leq C \sum_{j=1}^{\infty} j^{-\frac{2(r-\tau)}{d}} < \infty$$

which gives us the condition $\tau < -(d/2 - r)$. From here on we will assume that $\tau = r - s < r - d/2$.

Remark 3. Any elliptic operator $C_U \in \Psi^{-2r}$ that defines a non-negative symmetric operator $C_U : \mathcal{D}(N) = C^\infty(N) \rightarrow \mathcal{D}'(N)$ has the property that $B_U = C_U(I - \Delta)^r$ is in $\mathfrak{S}^1(H^\tau(N))$. By [3], these yield that C_U is a covariance operator of a random variable taking values in $H^\tau(N)$.

The operator B_U corresponds formally to the smoothness prior

$$\pi_{\text{pr}}(u) \underset{\text{formally}}{=} c \exp\left(-\frac{1}{2}(B_U^{-1}u, u)_{H^\tau}\right) = c \exp\left(-\frac{1}{2}\|C_U^{-1/2}u\|_{L^2}^2\right).$$

Notice that the realisations of U are almost surely not r times differentiable. In a case $r < d/2$ the realisations of U are almost surely not even in L^2 let alone differentiable. This is why we need to consider U as taking values in some space $H^\tau(N)$ with possible negative smoothness index τ .

4. Proof of the main theorems

Before we move to prove theorem 1 we will give a short introduction to hypoelliptic pseudodifferential operators.

Definition 3. Let $t, t_0 \in \mathbb{R}$. We define symbol class $HS^{-t, -t_0}(\mathbb{R}^d, \mathbb{R}^d)$ to consist of $a(x, \xi) \in C^\infty(\mathbb{R}^d, \mathbb{R}^d)$ for which

(1) For an arbitrary compact set $K \subset \mathbb{R}^d$ we can find such positive constants R, c_1 and c_2 that

$$c_1(1 + |\xi|)^{-t_0} \leq |a(x, \xi)| \leq c_2(1 + |\xi|)^{-t}, \quad |\xi| \geq R, x \in K.$$

(2) For any compact set $K \subset \mathbb{R}^d$ there exist constants R and $C_{\alpha, \beta, K}$ such that for all multi-indices α and β

$$|\partial_\xi^\alpha \partial_x^\beta a(x, \xi)| \leq C_{\alpha, \beta, K} |a(x, \xi)| (1 + |\xi|)^{-|\alpha|}, \quad |\xi| \geq R, x \in K.$$

We will denote by $H\Psi^{-t, -t_0}(N)$ the class of Ψ DO with local symbol $a(x, \xi) \in HS^{-t, -t_0}(V \times \mathbb{R}^d)$, see definition 2.

We denote $H^r(N) = H^r$ and $L^2(N) = L^2$ where N is a closed manifold and $\dim N = d$.

The proof of theorem 1 is rather long and technical so we will start by going through the main steps of it in a nutshell. The approximated solution we are studying is of the form

$$T_\delta(m_\delta) := \arg \min_{u \in H^r(N)} \{ \|Au\|_{L^2(N)}^2 - 2\langle m_\delta, Au \rangle + \delta^2 \|C_U^{-1/2}u\|_{L^2(N)}^2 \}.$$

As mentioned before the solution to this is

$$T_\delta(m_\delta) = (A^*A + \delta^2 C_U^{-1})^{-1} A^* m_\delta. \quad (4.1)$$

We can rewrite the above as

$$\begin{aligned} T_\delta(m_\delta) &= Z_\delta^{-1} A^* A u + Z_\delta^{-1} A^* (\varepsilon \delta) \\ &= u - \delta^2 Z_\delta^{-1} C_U^{-1} u + Z_\delta^{-1} A^* (\varepsilon \delta), \end{aligned} \quad (4.2)$$

where $Z_\delta = A^*A + \delta^2 C_U^{-1}$.

To study the convergence of the last term on the right-hand side of (4.2) we would like to write it in the form

$$Z_\delta^{-1} A^* (\varepsilon \delta) = \delta^{-1} F_\delta^{-1} (A^*A)^{-1} A^* \varepsilon, \quad (4.3)$$

where $F_\delta = (A^*A)^{-1} C_U^{-1} + \delta^{-2}$. In order to show that (4.3) is well-defined we first need to prove that A^*A and F_δ are invertible.

Lastly we study the converge of $\delta^2 Z_\delta^{-1} C_U^{-1} u$ and $\delta^{-1} F_\delta^{-1} (A^*A)^{-1} A^* \varepsilon$ to zero in appropriate Sobolev spaces and show that the latter term is always dominating.

We will start by showing that $A^*A \in H\Psi^{-2t, -2t_0}$ is invertible and $(A^*A)^{-1} \in H\Psi^{2t_0, 2t}$. Define $A^* : L^2(N) \rightarrow L^2(N)$ as the adjoint of an operator $A : L^2 \rightarrow L^2$. We assumed in theorem 1 that $A : L^2 \rightarrow L^2$ is one-to-one. Since $A^*A : H^r \rightarrow H^{r+2t}$, $r \in \mathbb{R}$, is hypoelliptic [49, propositions 5.2 and 5.3] $A^*A u = 0 \in C^\infty$ implies that $u \in C^\infty$ and hence $Au \in L^2$. Now we see that if $A^*A u = 0$ then

$$0 = (A^*A u, u)_{L^2} = (Au, Au)_{L^2} = \|Au\|_{L^2}^2$$

which implies $Au = 0$ and furthermore $u = 0$. Thus the operator $A^*A : H^r(N) \rightarrow H^{r+2t}(N)$ is one-to-one.

To study the mapping $A^*A \in H\Psi^{-2t, -2t_0}$ from some Sobolev space H^r , $r \in \mathbb{R}$, we define $L_r = A^*A : H^r \rightarrow H^{r+2t}$. The adjoint of L_r is denoted by $L_r' = (A^*A) : H^{-(r+2t)} \rightarrow H^{-r}$. Let $\phi, \psi \in C^\infty$. Then

$$\langle A^*A \phi, \psi \rangle_{H^{r+2t} \times H^{-(r+2t)}} = \langle \phi, A^*A \psi \rangle_{H^r \times H^{-r}},$$

that is, $L_r' = L_{-(r+2t)}$ and hence the adjoint is one-to-one. Now we can conclude that $L_r(H^r) \subset H^{r+2t}$ is a dense subset. Next we will prove two lemmas that show that the operator $A^*A : \mathcal{D}' \rightarrow \mathcal{D}'$ is also surjective.

Since $A^*A \in H\Psi^{-2t, -2t_0}$ is a hypoelliptic pseudodifferential operator it has a parametrix $B_1 \in H\Psi^{2t_0, 2t}$ [49, theorem 5.1]. Hence for any $r_0 > 0$ we get norm estimates

$$\begin{cases} \|A^*A u\|_{H^{r+2t}} \leq C_1 \|u\|_{H^r} \\ \|u\|_{H^r} \leq C_2 \|A^*A u\|_{H^{r+2t_0}} + C_3 \|u\|_{H^{r-r_0}}. \end{cases} \quad (4.4)$$

for all $u \in C^\infty$. Next we will show that C_3 is zero.

Lemma 1. *Let $A^*A \in H\Psi^{-2t, -2t_0}$ be an injective hypoelliptic pseudodifferential operator. Then we have the following estimates*

$$C_1 \|A^*Au\|_{H^{r+2t}} \leq \|u\|_{H^r} \leq C_2 \|A^*Au\|_{H^{r+2t_0}}.$$

Proof. We get the first inequality since A^*A is continuous linear operator. If the second estimate in (4.4) is not valid with $C_3 = 0$ for any $C_2 > 0$ then we can choose a sequence u_j such that $\|u_j\|_{H^{r-r_0}} = 1$ and $\|u_j\|_{H^r} \geq j \|A^*Au_j\|_{H^{r+2t_0}}$. When $j > 2C_2$ then

$$\begin{aligned} \|u_j\|_{H^r} &\leq C_2 \|A^*Au_j\|_{H^{r+2t_0}} + C_3 \|u_j\|_{H^{r-r_0}} \\ &\leq \frac{1}{2} \|u_j\|_{H^r} + C_3 \|u_j\|_{H^{r-r_0}}. \end{aligned}$$

This gives us

$$\|u_j\|_{H^r} \leq 2C_3 \|u_j\|_{H^{r-r_0}} = 2C_3.$$

Since $r_0 > 0$, the embedding $H^r \hookrightarrow H^{r-r_0}$ is compact. Now there exists a subsequence u_{j_ℓ} and such a $w \in H^{r-r_0}$ that $\lim_{\ell \rightarrow \infty} u_{j_\ell} = w$ in H^{r-r_0} . We assumed that $\|u_{j_\ell}\|_{H^{r-r_0}} = 1$ which implies $\|w\|_{H^{r-r_0}} = 1$. On the other hand

$$j_\ell \|A^*Au_{j_\ell}\|_{H^{r+2t_0}} \leq \|u_{j_\ell}\|_{H^r} \leq 2C_3,$$

that is, $\lim_{\ell \rightarrow \infty} \|A^*Au_{j_\ell}\|_{H^{r+2t_0}} = 0$ and because $-r_0 + 2t \leq 2t_0$

$$\lim_{\ell \rightarrow \infty} \|A^*Au_{j_\ell}\|_{H^{r-r_0+2t}} = 0.$$

Since $u_{j_\ell} \rightarrow w$ in H^{r-r_0} we also have $A^*Au_{j_\ell} \rightarrow A^*Aw$ in H^{r-r_0+2t} . Combining the above results we see that $\|A^*Aw\|_{H^{r-r_0+2t}} = 0$. Operator A^*A is one to one and hence $w = 0$. This is a contradiction since $\|w\|_{H^{r-r_0}} = 1$. \square

Lemma 2. Let $A^*A : H^r \rightarrow H^{r+2t}$ be an injective hypoelliptic operator. Then the image of H^r in the map A^*A satisfies

$$H^{r+2t_0} \subset A^*A(H^r) \subset H^{r+2t}.$$

Proof. The second inclusion is a direct consequence of the mapping properties of A^*A . Let $f \in H^{r+2t_0}$. Since $C^\infty \subset H^{r+2t_0}$ is a dense subset we can find such a sequence $f_j \in C^\infty$ that $\lim_{j \rightarrow \infty} f_j = f$ in H^{r+2t_0} . Since $A^*A(H^r) \subset H^{r+2t}$ is dense we can also choose a sequence $h_{j,\ell} = A^*Ag_{j,\ell} \in A^*A(H^r)$ such that $\lim_{\ell \rightarrow \infty} h_{j,\ell} = f_j$ in H^{r+2t_0} . Denote $g_j = g_{j,\ell_j} \in H^r$ for which $\lim_{j \rightarrow \infty} A^*Ag_j = f$ in H^{r+2t_0} . Using lemma 1 we see

$$\lim_{j,k \rightarrow \infty} \|g_j - g_k\|_{H^r} \leq C_2 \lim_{j,k \rightarrow \infty} \|A^*Ag_j - A^*Ag_k\|_{H^{r+2t_0}} = 0.$$

Hence also $g_j \in H^r$ is a Cauchy sequence. Thus there exists such $g \in H^r$ that $\lim_{j \rightarrow \infty} g_j = g$ in H^r . On the other hand

$$\lim_{j \rightarrow \infty} \|A^*Ag_j - A^*Ag\|_{H^{r+2t}} \leq C_1 \lim_{k \rightarrow \infty} \|g_j - g\|_{H^r} = 0.$$

Combining the above we get $A^*Ag = f$. \square

Using lemma 2 we see that $A^*A(\mathcal{D}') = \mathcal{D}'$, that is the operator A^*A is also onto. Now we can conclude that there exists an inverse operator $(A^*A)^{-1} : \mathcal{D}' \rightarrow \mathcal{D}'$. It remains to show that the inverse operator is a hypoelliptic pseudodifferential operator.

Lemma 3. *A self-adjoint, smoothing, one-to-one hypoelliptic operator $A^*A \in H\Psi^{-2t, -2t_0}$ has an inverse operator $(A^*A)^{-1} \in H\Psi^{2t_0, 2t}$.*

Proof. Denote $B = (A^*A)^{-1} : \mathcal{D}' \rightarrow \mathcal{D}'$. For an operator $A^*A : H^r \rightarrow H^{r+2t}$ we define $B_0 \subset B$ with domain

$$\mathcal{D}(B_0) = \{f \in H^{r+2t} \mid Bf \in H^r\} = A^*A(H^r).$$

Using the hypoellipticity of A^*A we see that $A^*Au = f \in C^\infty$ implies $u \in C^\infty$. This gives us $B_0 : C^\infty \rightarrow C^\infty$. Since C^∞ is a Frechét space and A^*A is continuous and linear $A^*A : C^\infty \rightarrow C^\infty$ is an open mapping [46, theorem 2.11]. Hence the operator $B_0 : C^\infty \rightarrow C^\infty$ is continuous.

Since A^*A is hypoelliptic it has a parametrix $B_1 \in H\Psi^{2t_0, 2t}$ [49, theorem 5.1]

$$\begin{cases} B_1(A^*A) = I + K_1, & K_1 \in \Psi^{-\infty} \\ (A^*A)B_1 = I + K_2, & K_2 \in \Psi^{-\infty} \end{cases}$$

and we can write

$$B_0 = B_0((A^*A)B_1 - K_2) = B_1 - B_0K_2.$$

The operator $B_0K_2 : \mathcal{D}' \rightarrow C^\infty$ is continuous and thus we have shown that

$$B_0 = B_1 \text{ mod } \Psi^{-\infty}.$$

That is $B_0 \in H\Psi^{2t_0, 2t}$. □

Next we will examine Ψ DOs that depend on spectral variable $\lambda = \delta^{-2}$. For the general theory see [49].

Definition 4. The symbol class $S_p^m(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}_+)$ consist of the functions $a(x, \xi, \lambda)$ such that

- (1) $a(x, \xi, \lambda_0) \in C^\infty(\mathbb{R}^d \times \mathbb{R}^d)$ for every fixed $\lambda_0 \geq 0$ and
- (2) for arbitrary multi-indices α and β and any compact set $K \subset \mathbb{R}^d$ there exists a constant $C_{\alpha, \beta, K}$ such that

$$|\partial_\xi^\alpha \partial_x^\beta a(x, \xi, \lambda)| \leq C_{\alpha, \beta, K} (1 + |\xi| + |\lambda|^{1/p})^{m-|\alpha|}$$

for $x \in K$, $\xi \in \mathbb{R}^d$ and $\lambda \geq 0$.

We denote by $\Psi_p^m(N, \mathbb{R}_+)$ the class of pseudodifferential operators A_λ for which the local symbol $a(x, \xi, \lambda) \in S_p^m(V \times \mathbb{R}^d, \mathbb{R}_+)$, see definition 2.

Definition 5. If there are constants $C_1, C_2, R > 0$ such that the symbol $a(x, \xi, \lambda) \in S_p^m(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}_+)$ satisfies

$$C_1(|\xi| + |\lambda|^{1/p})^{m_0} \leq |a(x, \xi, \lambda)| \leq C_2(|\xi| + |\lambda|^{1/p})^m,$$

for $|\xi| + |\lambda| \geq R$, we say that a is hypoelliptic with parameter λ and denote $a(x, \xi, \lambda) \in HS_p^{m, m_0}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}_+)$. We will denote by $H\Psi_p^{m, m_0}(N, \mathbb{R}_+)$ the class of Ψ DOs depending on the parameter λ whose local symbol belongs to $HS_p^{m, m_0}(V \times \mathbb{R}^d, \mathbb{R}_+)$, see definition 2.

Next we will prove that

$$F_\lambda = (A^*A)^{-1}C_U^{-1} + \lambda$$

is invertible. Operator $F_\lambda \in H\Psi^{2(t_0+r), 2(t+r)}(N)$ is hypoelliptic since $(A^*A)^{-1}C_U^{-1} \in H\Psi^{2(t_0+r), 2(t+r)}(N)$ is hypoelliptic and $\lambda I \in \Psi^0(N)$. Denote $Q = (A^*A)^{-1}C_U^{-1}$ and its symbol $q(x, \xi) \in HS^{2(t_0+r), 2(t+r)}(N)$. Then for the symbol $\sigma(F_\lambda)(x, \xi) = q(x, \xi) + \lambda$ of the operator F_λ

$$|\partial_\xi^\alpha \partial_x^\beta (q(x, \xi) + \lambda)| \leq C_{\alpha, \beta} (1 + |\xi| + |\lambda|^{1/(2(t_0+r))})^{2(t_0+r) - |\alpha|}.$$

By [49, theorem 9.2.] there exist $R > 0$ such that for $|\lambda| \in [R, \infty)$ the operator $F_\lambda \in H\Psi_{2(t_0+r)}^{2(t_0+r), 2(t+r)}(N, \mathbb{R}_+)$ is invertible with

$$F_\lambda^{-1} \in H\Psi_{2(t_0+r)}^{-2(t+r), -2(t_0+r)}(N, [R, \infty)).$$

We have now shown that the operator Z_δ can be written

$$Z_\delta^{-1} = \lambda((A^*A)^{-1}C_U^{-1} + \lambda I)^{-1}(A^*A)^{-1}, \quad (4.5)$$

where $\lambda = \delta^{-2}$. Hence we can rewrite (4.2)

$$T_\lambda(m) = u - \lambda^{-1}Z_\delta^{-1}C_U^{-1}u + \sqrt{\lambda}F_\lambda^{-1}(A^*A)^{-1}A^*\varepsilon. \quad (4.6)$$

Now we will proceed to study the convergence of the second and third term on the right-hand side of (4.6). For the third term of (4.6) we have $(A^*A)^{-1}A^* : H^{-s} \rightarrow H^k$, $k = -s + t - 2t_0$ and $F_\lambda^{-1} : H^k \rightarrow H^{k+2(t+r)}$. Hence when $\zeta \leq k + 2(t+r)$ we have

$$\|F_\lambda^{-1}(A^*A)^{-1}A^*\varepsilon\|_{H^\zeta} \leq \|F_\lambda^{-1}\|_{k, \zeta} \|(A^*A)^{-1}A^*\varepsilon\|_{H^k},$$

where $\|F_\lambda^{-1}\|_{k, \zeta}$ is the norm of $F_\lambda^{-1} : H^k(N) \rightarrow H^\zeta(N)$ and $k, \zeta \in \mathbb{R}$. Next we want to study what happens to the norm when $\lambda \rightarrow \infty$.

We have the following norm estimates for $F_\lambda^{-1} \in \Psi_p^m(N, \mathbb{R}_+)$ when $\ell \geq m$ and λ large enough [49, theorem 9.1]

$$\|F_\lambda^{-1}\|_{k, k-\ell} \leq C_{k, \ell} (1 + |\lambda|^{1/p})^m, \quad \text{if } \ell \geq 0, \quad (4.7)$$

$$\|F_\lambda^{-1}\|_{k, k-\ell} \leq C_{k, \ell} (1 + |\lambda|^{1/p})^{-(\ell-m)}, \quad \text{if } \ell \leq 0. \quad (4.8)$$

In our case $F_\lambda^{-1} \in \Psi_p^m$ where $m = -2(t+r)$ and $p = 2(t_0+r)$. We will write $\|F_\lambda^{-1}\|_{k, \zeta} = \|F_\lambda^{-1}\|_{k, k-\ell}$ where $\ell = k - \zeta \geq m$.

First we study the case when $\ell \geq 0$ that is $\zeta \leq k$. Inequality (4.7) gives us the norm estimate

$$\|F_\lambda^{-1}\|_{k, k-\ell} \leq C(1 + |\lambda|^{1/p})^m.$$

Because we want $\sqrt{\lambda}\|F_\lambda^{-1}(A^*A)^{-1}A^*\varepsilon\|_{H^\zeta}$ to converge when $\lambda \rightarrow \infty$ we have to require that

$$\frac{m}{p} = \frac{-2(t+r)}{2(t_0+r)} < -\frac{1}{2}.$$

This is true when $t_0 < 2t+r$.

When $\ell \leq 0$ we have $k \leq \zeta \leq k + 2(t+r)$ and can use (4.8)

$$\|F_\lambda^{-1}\|_{k, k-\ell} \leq C(1 + |\lambda|^{1/p})^{-(\ell-m)}.$$

For convergence we need

$$\frac{m - \ell}{p} = \frac{-2(t + r) - k + \zeta}{2(t_0 + r)} < -\frac{1}{2}$$

that is $k \leq \zeta < k + 2t + r - t_0 = r - s - 3(t_0 - t)$ which can be true only if $t_0 < 2t + r$.

Next we will prove the convergence of the term $\delta^2 Z_\delta^{-1} C_U^{-1} u$ in H^ζ . Since we got above that $\zeta < \tau - 3(t_0 - t)$ we can write $\zeta = \tau - \theta$ where $\theta \geq 3(t_0 - t) \geq 0$. We need to find such $\eta \geq 0$ and $\gamma \geq 0$ that $\gamma + \eta = 1$ and $t_0\gamma - r\eta + r - \theta/2 = 0$. Define $\gamma = \theta/2(t_0 + r)$ and $\eta = 1 - \theta/2(t_0 + r)$. Now $\eta \geq 0$ only if $\theta \leq 2(t_0 + r)$. Hence we will choose $\theta = \min\{\tau - \zeta, 2(t_0 + r)\}$.

Since $Z_\delta = A^*A + \delta^2 C_U^{-1}$ where $A^*A \geq c_1(I - \Delta)^{-t_0}$ and $c_2(I - \Delta)^r \leq C_U^{-1} \leq c_3(I - \Delta)^r$ we get

$$\begin{aligned} \|\delta^2 Z_\delta^{-1} C_U^{-1} u\|_{H^\zeta} &\leq \delta^2 \|(A^*A)^{-\gamma} (\delta^2 C_U^{-1})^{-\eta} c_3 (I - \Delta)^{r+\frac{\zeta}{2}} u\|_{L^2} \\ &\leq \delta^2 \|(c_1(I - \Delta)^{-t_0})^{-\gamma} (c_2 \delta^2 (I - \Delta)^r)^{-\eta} (I - \Delta)^{r-\frac{\theta}{2}+\frac{\zeta}{2}} u\|_{L^2} \\ &= C \delta^{\frac{\theta}{2(t_0+r)}} \|u\|_{H^\tau}, \end{aligned} \tag{4.9}$$

where $\theta = \min\{\tau - \zeta, 2(t_0 + r)\}$.

Adding the above results together we can prove theorem 1.

Proof of theorem 1. To get the speed of convergence we use the fact that U and \mathcal{E} are independent. Similarly to (4.6) we get

$$U_\delta(\omega) - U(\omega) = -\delta^2 Z_\delta^{-1} C_U^{-1} U(\omega) + \frac{1}{\delta} F_{\delta^{-2}}^{-1} (A^*A)^{-1} A^* \mathcal{E}(\omega), \tag{4.10}$$

where by (4.9),

$$\|\delta^2 Z_\delta^{-1} C_U^{-1} U\|_{H^\zeta} \leq C \delta^{\frac{\theta}{2(t_0+r)}} \|U\|_{H^\tau}. \tag{4.11}$$

For the second part on the right-hand side of (4.10) we can write

$$\mathbb{E} \left\| \frac{1}{\delta} F_{\delta^{-2}}^{-1} (A^*A)^{-1} A^* \mathcal{E}(\omega) \right\|_{H^\zeta}^p \leq \left\| \frac{1}{\delta} F_{\delta^{-2}}^{-1} (A^*A)^{-1} A^* \right\|_{H^{-s} \rightarrow H^\zeta}^p \mathbb{E} \|\mathcal{E}(\omega)\|_{H^{-s}}^p, \tag{4.12}$$

with $p \in \{1, 2\}$ and $\theta = \min\{\tau - \zeta, 2(t_0 + r)\}$.

When $\zeta \leq t - s - 2t_0$ we get

$$\begin{aligned} \mathbb{E} \|U_\delta(\omega) - U(\omega)\|_{H^\zeta} &\leq \delta^2 \mathbb{E} \|Z_\delta^{-1} C_U^{-1} U(\omega)\|_{H^\zeta} + \delta^{-1} \mathbb{E} \|F_\delta^{-1} (A^*A)^{-1} A^* \mathcal{E}(\omega)\|_{H^\zeta} \\ &\leq C_1 \delta^{\frac{\theta}{2(t_0+r)}} \mathbb{E} \|U(\omega)\|_{H^\tau} + C_2 \delta^{-1+\frac{2(t+r)}{2(t_0+r)}} \mathbb{E} \|\mathcal{E}\|_{H^{-s}} \\ &\leq C \max\{\delta^{\frac{\theta}{2(t_0+r)}}, \delta^{\frac{2t+r-t_0}{2(t_0+r)}}\}, \end{aligned}$$

where $\theta = \min\{\tau - \zeta, 2(t_0 + r)\}$. Next we will study which of the terms is dominating. The noise term $\delta^{-1} F_\delta^{-1} (A^*A)^{-1} A^* \mathcal{E}(\omega)$ is dominating if

$$2t + r - t_0 \leq \theta.$$

Assume first that $\theta = \tau - \zeta$. Then

$$\theta = \tau - \zeta \geq 2t_0 + r - t \geq 2t + r - t_0.$$

If $\theta = 2(t_0 + r)$ we get

$$\theta = 2(t_0 + r) \geq 2t_0 + t_0 - 2t + r \geq 2t + r - t_0$$

since $t \leq t_0 < 2t + r$. Hence the noise term is dominating in both cases and we have proven

$$\mathbb{E}\|U_\delta(\omega) - U(\omega)\|_{H^\zeta} \leq C\delta^{\frac{2t-t_0+r}{t_0+r}}.$$

If $t - s - 2t_0 \leq \zeta < \tau - 3(t_0 - t)$ we get

$$\begin{aligned} \mathbb{E}\|U_\delta(\omega) - U(\omega)\|_{H^\zeta} &\leq \delta^2 \mathbb{E}\|Z_\delta^{-1}C_U^{-1}U(\omega)\|_{H^\zeta} + \delta^{-1} \mathbb{E}\|F_\delta^{-1}(A^*A)^{-1}A^*\mathcal{E}(\omega)\|_{H^\zeta} \\ &\leq C_1 \delta^{\frac{\tau-\zeta}{t_0+r}} \mathbb{E}\|U(\omega)\|_{H^\tau} + C_2 \delta^{-1+\frac{2(t+r)+\ell}{t_0+r}} \mathbb{E}\|\mathcal{E}(\omega)\|_{H^{-s}} \\ &\leq C \max \left\{ \delta^{\frac{\tau-\zeta}{t_0+r}}, \delta^{\frac{2t+r-t_0+\ell}{t_0+r}} \right\}. \end{aligned}$$

Above $\ell = t - s - 2t_0 - \zeta$. Note that when $\zeta \geq t - s - 2t_0$ then $\tau - \zeta \leq 2(r + t_0)$. The noise term is dominating if

$$2t + r - t_0 + \ell \leq \tau - \zeta.$$

This is always true since $\tau = r - s$ and

$$2t + r - t_0 + \ell = r - s - \zeta - 3(t_0 - t) \leq \tau - \zeta.$$

Hence we can conclude

$$\mathbb{E}\|U_\delta(\omega) - U(\omega)\|_{H^\zeta} \leq C\delta^{\frac{2t+r-t_0+\ell}{t_0+r}}.$$

□

Proof of theorem 2. Similarly to (4.10) we get

$$U_\delta^\dagger(\omega) - u^\dagger = -\delta^2 Z_\delta^{-1} C_U^{-1} u^\dagger + \frac{1}{\delta} F_{\delta^{-2}}^{-1} (A^*A)^{-1} A^* \mathcal{E}(\omega), \quad (4.13)$$

where the first term on the right side satisfies, by (4.9),

$$\|\delta^2 Z_\delta^{-1} C_U^{-1} u^\dagger\|_{H^\zeta} \leq C\delta^{\frac{\theta}{t_0+r}} \|u^\dagger\|_{H^\tau} \quad (4.14)$$

with $\theta = \min\{\tau - \zeta, 2(t_0 + r)\}$. The expectation of the second term in the right side is estimated in (4.12). Analysing the obtained terms as in the proof of theorem 1, we obtain

$$\mathbb{E}\|U_\delta^\dagger(\omega) - u^\dagger\|_{H^\zeta}^2 \leq C(1 + \|u^\dagger\|_{H^\tau}^2) \delta^{\frac{2(\tau-3(t_0-\ell))}{t_0+r}}.$$

□

5. Posterior distribution and confidence regions

One advantage Bayesian inversion offers over deterministic regularisation is uncertainty quantification. Since the solution to the Bayesian inverse problem is the posterior distribution of the unknown we can study its credible sets and their contraction in some Sobolev space H^ζ when $\delta \rightarrow 0$. A Bayesian credible set is a region in the posterior distribution that contains a large fraction of the posterior mass, for instance, 95%. We are dealing with Gaussian distributions so we define the credible sets to be central regions. This means these sets are defined as central balls with u_δ as a centre.

The above mentioned credible sets are often used to visualise the remaining Bayesian uncertainty in the estimate. Frequentists use another kind of uncertainty quantification called confidence region. A confidence region is a range of values that frequently includes the unknown of interest if the experiment is repeated. We can define confidence regions as central balls with u_δ^\dagger as the centre. Here u_δ^\dagger is the frequentist approximated solution generated by a

true solution u^\dagger . How frequently the ball around the approximated solution, with different realisation of the noise, contains the true solution is determined by the confidence level. See for example [15, 56].

In the finite-dimensional parametric case and under mild conditions on the prior Bernstein–von Mises theorem provides that the credible sets of smooth models are asymptotically equivalent with the frequentist confidence regions based on the maximum likelihood estimator, see [56]. In an infinite-dimensional case there is no corresponding theorem and Bayesian credible sets are not automatically frequentist confidence sets. This means that if we assume that the data is generated by a ‘true parameter’, it is not automatically true that credible sets contain that truth with probability at least the credible level. However the correspondence of Bayesian and frequentist uncertainty has been studied in many recent papers see e.g. [4, 5, 27, 35, 43, 52]. These results are important since they show that some credible sets can give a good idea of the uncertainty of the estimate in the classical sense. In this section we show that the posterior distribution converges and we give some convergence rates. We also prove that in the elliptic case the frequentist posterior contractions rates agrees, up to $\varepsilon > 0$ arbitrarily small, with the minimax convergence rate. We do not address the question about the frequentist coverage of the credible sets.

We will start by studying the convergence of the posterior covariance C_δ which, with the convergence of the posterior mean U_δ , guarantees the convergence of the posterior distribution.

When $U \sim N(0, C_U)$, $\mathcal{E} \sim N(0, I)$ and

$$M_\delta = AU + \delta\mathcal{E} \quad (5.1)$$

the conditional probability distribution of U with respect to the measurement M_δ is a Gaussian measure with mean U_δ and covariance [36, 37]

$$C_\delta = C_U - C_U A^* (A C_U A^* + \delta^2 I)^{-1} A C_U. \quad (5.2)$$

If $A^* : \mathcal{D}'(N) \rightarrow \mathcal{D}'(N)$ is invertible we can rewrite the above

$$\begin{aligned} C_\delta &= \delta^2 (A^* A + \delta^2 C_U^{-1})^{-1} \\ &= ((A^* A)^{-1} C_U^{-1} + \delta^{-2} I)^{-1} (A^* A)^{-1}. \end{aligned} \quad (5.3)$$

Note that the covariance operator is deterministic and thus independent of M_δ .

We define $F_\lambda = (A^* A)^{-1} C_U^{-1} + \lambda I$, where $\lambda = \delta^{-2}$, as in section 4. Then

$$F_\lambda^{-1} \in \Psi_p^m(N, [R, \infty)),$$

where $m = -2(t + r)$ and $p = 2(t_0 + r)$. Using the norm estimate (4.8) we get

$$\begin{aligned} \|\delta^2 Z_\delta^{-1}\|_{-\tau, \tau} &= \|((A^* A)^{-1} C_U^{-1} + \delta^{-2} I)^{-1}\|_{-\tau-2t, \tau} \|(A^* A)^{-1}\|_{-\tau, -\tau-2t} \\ &\leq c \|F_\lambda^{-1}\|_{k, k-\ell} \\ &\leq c (1 + \lambda^{\frac{1}{p}})^{-(\ell-m)}. \end{aligned}$$

Above $k = -\tau - 2t$ and $\ell = -2(\tau + t) < 0$. Since $\tau = r - s$ we can write

$$\frac{\ell - m}{p} = \frac{-\tau + r}{t_0 + r} = \frac{s}{t_0 + r}$$

and hence we get the following convergence rate for the posterior covariance

$$\|C_\delta\|_{-\tau, \tau} \leq c \delta^{\frac{2s}{t_0+r}} \leq c \delta^{\frac{d}{t_0+r}}. \quad (5.4)$$

We see that the more smoothing the forward operator A is the worse convergence we get. Note that r and s do not only affect the convergence speed but also the spaces between which the norm is taken.

Remark 4. Observe that the random variable U takes values in H^τ and the estimate (5.4) concerns the mapping properties of the posterior covariance operator C_δ from the dual space $H^{-\tau} = (H^\tau)'$ to the space H^τ . For strictly positive $\delta > 0$ the MAP estimator U_δ belongs to the space H^r , $r = \tau + s \geq \tau$, but as $\delta \rightarrow 0$, the MAP estimators U_δ converge in a less regular space H^ζ , $\zeta < \tau - 3(t_0 - t) \leq \tau$, see (2.10).

5.1. Contraction of the posterior distribution

Next we consider the inverse problem using the frequentist setting described in section 2.1.2 with the additional assumption that $\tau > 0$. We assume below that C_U satisfies the assumptions in theorem 2 that in particular imply that C_U is the covariance operator of a random variable U taking values in $H^\tau(N)$, see remark 3. We recall that we consider a fixed ‘true’ solution $u^\dagger \in H^\tau(N)$ and the noise model $M_\delta^\dagger(\omega) = Au^\dagger + \delta\mathcal{E}(\omega)$ as in (2.14). Also, note that the MAP-estimate is then $U_\delta^\dagger = T_\delta(M_\delta^\dagger)$.

In the frequentist case one is often interested in the limiting behaviour of the posterior measure $P_{M_\delta^\dagger}$ when $\delta \rightarrow 0$. Here, $P_{M_\delta^\dagger}$ is a random measure in $H^\tau(N)$, depending on δ and the MAP estimator $U_\delta^\dagger = U_\delta^\dagger(\omega) = T_\delta(M_\delta^\dagger(\omega))$ (that further depends on the deterministic variable u^\dagger and the realisation $\mathcal{E}(\omega)$ of the random noise). Let W_δ be a Gaussian random variable, taking values in $H^\tau(N)$, that is independent of the noise \mathcal{E} , has zero mean and the covariance operator C_δ , see (3.4). For a measurable set $B \subset H^\tau(N)$ we define

$$P_{M_\delta^\dagger}(B) = \mu_\delta(\{w_\delta \in H^\tau(N) \mid w_\delta = b - U_\delta^\dagger(\omega) \text{ with } b \in B\}), \quad (5.5)$$

where $\mu_\delta = N(0, C_\delta)$. Roughly speaking, $P_{M_\delta^\dagger}$ is a Gaussian measure in $H^\tau(N)$ with the mean U_δ^\dagger and the covariance operator C_δ .

Recall that we consider the probability space $(\Omega, \Sigma, \mathbb{P})$ and denote by χ_S the indicator function of S . Let $S \in \Sigma$. We use the notations

$$\mathbb{E}_{M_\delta^\dagger}(F(\omega, u^\dagger)) := \mathbb{E}(F(\omega, u^\dagger)|\mathcal{F}), \quad \mathbb{P}_{M_\delta^\dagger}(S) := \mathbb{P}(S|\mathcal{F}) = \mathbb{E}(\chi_S|\mathcal{F}), \quad (5.6)$$

for the conditional expectation and conditional probability. Above $\mathcal{F} \subset \Sigma$ is the σ -algebra generated by random variable $M_\delta^\dagger(\omega)$ or equivalently, the noise $\mathcal{E}(\omega)$. Roughly speaking, in the notation $\mathbb{E}_{u^\dagger} F$ the subindex u^\dagger reminds that u^\dagger is a fixed parameter and the expectation is taken only with respect the noise. The notation $P_{M_\delta^\dagger}(B)$ indicates that the measure of B is computed using the posterior probability measure which mean $U_\delta^\dagger = T_\delta(M_\delta^\dagger)$ depends on the measurement M_δ^\dagger . Since the random variable W_δ has distribution μ_δ , we have by (5.5)

$$P_{M_\delta^\dagger}(B) = \int_{H^\tau(N)} \chi_B(w_\delta + U_\delta^\dagger) d\mu_\delta(w_\delta) = \mathbb{P}(\{W_\delta + U_\delta^\dagger \in B\}). \quad (5.7)$$

Following the approach in [11, 14, 27, 57] we next show that the posterior measure contracts to a Dirac measure centred on the fixed true solution u^\dagger .

Theorem 3. *Let $r > s > d/2$ and N be a d -dimensional closed manifold. Let $u^\dagger \in H^\tau(N)$ where $\tau = r - s > 0$. Assume that C_U , the covariance operator of the Gaussian prior, is a self-adjoint, injective and elliptic pseudodifferential operator of order $-2r$. Let $\mathcal{E}(y, \omega)$ be white Gaussian noise on N . Consider the measurement*

$$M_\delta^\dagger(y, \omega) = A(u^\dagger(\cdot)) + \delta\mathcal{E}(y, \omega), \quad \omega \in \Omega,$$

where $A \in H\Psi^{-t, -t_0}$, $t > \max\{0, -\tau + s\}$ and $t \leq t_0 \leq t + \tau/3$, is a hypoelliptic pseudodifferential operator on the manifold N and $A : L^2(N) \rightarrow L^2(N)$ is injective. We assume also that $A^* : \mathcal{D}'(N) \rightarrow \mathcal{D}'(N)$ is invertible. Above $\delta \in \mathbb{R}_+$ is the noise level and \mathcal{E} takes values in $H^{-s}(N)$ with some $s > d/2$. Let U_δ^\dagger be the MAP estimated given by (2.15).

Let $\kappa < \kappa_0 = \frac{2(\tau - 3(t_0 - t))}{t_0 + r}$, $c_0 > 0$, and $R > 0$. Then there is $c_1 > 0$ such that

$$\sup_{\|u^\dagger\|_{H^r} \leq R} \mathbb{E}_{u^\dagger} P_{M_\delta^\dagger} \{u \in H^r(N) \mid \|u - u^\dagger\|_{L^2(N)} \geq c_0 \delta^\kappa\} \leq c_1 \delta^{2(\kappa_0 - \kappa)} \rightarrow 0, \quad (5.8)$$

as $\delta \rightarrow 0$.

Proof. Let $u^\dagger \in H^r(N)$ and W_δ be the Gaussian variable defined above. Using the Markov inequality and (5.7), we get

$$\begin{aligned} P_{M_\delta^\dagger} \{u \in H^r(N) \mid \|u - u^\dagger\|_{L^2} \geq c_0 \delta^\kappa\} &= \mathbb{P}_{M_\delta^\dagger}(\{\|W_\delta + U_\delta^\dagger - u^\dagger\|_{L^2(N)} \geq c_0 \delta^\kappa\}) \\ &\leq \frac{1}{(c_0 \delta^\kappa)^2} \mathbb{E}_{M_\delta^\dagger}(\|W_\delta + U_\delta^\dagger - u^\dagger\|_{L^2(N)}^2). \end{aligned} \quad (5.9)$$

Since W_δ and U_δ^\dagger are independent and W_δ has the covariance operator C_δ , we obtain using notations (5.6)

$$\begin{aligned} \mathbb{E}_{u^\dagger} \mathbb{E}_{M_\delta^\dagger}(\|W_\delta + U_\delta^\dagger - u^\dagger\|_{L^2(N)}^2) &= \mathbb{E}_{u^\dagger} \mathbb{E}_{M_\delta^\dagger} \|W_\delta\|_{L^2}^2 + \mathbb{E}_{u^\dagger} \mathbb{E}_{M_\delta^\dagger} \|U_\delta^\dagger - u^\dagger\|_{L^2}^2 \\ &= \text{Tr}_{L^2(N) \rightarrow L^2(N)}(C_\delta) + \mathbb{E}_{u^\dagger} \|U_\delta^\dagger - u^\dagger\|_{L^2(N)}^2. \end{aligned} \quad (5.10)$$

We have shown in theorem 2 that the second term on the right side of (5.10) can be estimated by $c_2(1 + R^2)\delta^{2\kappa_0}$ with some $c_2 > 0$. Hence it is enough to show that

$$\text{Tr}_{L^2 \rightarrow L^2}(C_\delta) \leq c\delta^{2\kappa_0}$$

with $\kappa_0 = \frac{2(\tau - 3(t_0 - t))}{t_0 + r}$. We can estimate the trace by writing

$$\begin{aligned} \text{Tr}_{L^2 \rightarrow L^2}(C_\delta) &= \text{Tr}_{L^2 \rightarrow L^2}((I - \Delta)^{-s}(I - \Delta)^s C_\delta) \\ &\leq \text{Tr}_{L^2 \rightarrow L^2}((I - \Delta)^{-s}) \|(I - \Delta)^s C_\delta\|_{L^2 \rightarrow L^2}. \end{aligned}$$

Above $(I - \Delta)^{-s}$ is trace class operator in L^2 since

$$\sum_{j=1}^{\infty} (j^{2/d})^{-s} < \infty \quad \text{when } s > d/2.$$

As before we get

$$\begin{aligned} \|(I - \Delta)^s C_\delta\|_{0,0} &= \|(I - \Delta)^s\|_{2,s,0} \|\mathcal{F}_\lambda^{-1}\|_{-2t,2s} \|(A^* A)^{-1}\|_{0,-2t} \\ &\leq c \|\mathcal{F}_\lambda^{-1}\|_{-2t,-2t-\ell} \leq c(1 + \lambda^{\frac{1}{p}})^{-(\ell-m)}. \end{aligned}$$

Above $\ell = -2(s + t) < 0$. We can write

$$\frac{\ell - m}{p} = \frac{\tau}{t_0 + r}$$

and hence

$$\text{Tr}_{L^2 \rightarrow L^2}(C_\delta) \leq c\delta^{\frac{2\tau}{t_0+r}} \leq c\delta^{2\kappa_0}.$$

□

Note that in the elliptic case $t_0 = t$ we get contraction

$$\mathbb{E}_{u^\dagger} P_{M_\delta^\dagger} \{u \in H^\tau(N) \mid \|u - u^\dagger\|_{L^2(N)} \geq c_0 \delta^\kappa\} \rightarrow 0$$

when $\delta \rightarrow 0$ for all $c_0 > 0$ and $\kappa < \frac{2\tau}{s+\tau+t}$. Since $s = \frac{d}{2} + \epsilon$ the above convergence rate agrees, up to $\epsilon > 0$ arbitrarily small, with the minimax convergence rate.

Remark 5. Above we have assumed that u^\dagger is in H^τ . This corresponds to the fact that the random variable U , having the covariance operator $C_U \in \Psi^{-2r}$, takes values in H^τ . The $L^2(N)$ norm in the contraction formula (5.8) can be considered as a loss function on $H^\tau(N)$. Note that the loss function $d(v_1, v_2) = \|v_1 - v_2\|_{L^2}$ defines a distance function in the $H^\tau(N)$, but the obtained metric space is not complete. When the direct map A is the identity map, similar estimates with different loss functions have been studied in a general setting in [18]. However, from the point of view of inverse problems [18] corresponds to the case when the direct operator and the covariance operator of the prior commute. This differs from the problem analysed in our paper, where covariance operator C_U and the operator A may not commute, and are of quite different type in the sense that C_U is an elliptic operator but A is hypoelliptic operator. The phenomenon that the solution u^\dagger is assumed to be in a smoother space, in our case in H^τ , and the convergence of the posterior distribution is analysed using a loss function given by a less strict norm, in our case L^2 -norm, appears in many frequentist studies, see e.g. theorems 2.2 and 2.3 and remark 3.6 in [1]. Conditions similar to the smoothness requirement $u^\dagger \in H^\tau$ are also encountered in classical regularisation theory [9] where this type of conditions are called source conditions.

5.2. Convergence of the posterior distribution in Bayesian settings

Next we will proceed to study the contraction of the posterior distribution using Bayesian techniques, the measurement model $M_\delta = AU + \delta\mathcal{E}$ and the MAP estimator $U_\delta = T_\delta(M_\delta)$. Let us write

$$V_\delta = U_\delta + W_\delta,$$

where $W_\delta \sim N(0, C_\delta)$ is a Gaussian variable having the covariance operator C_δ given in (5.3) and the zero mean. Random variables W_δ and U_δ are assumed to be independent. Let $\mathcal{M}_\delta = \sigma(M_\delta)$ be the σ -algebra generated by the random variable M_δ . Then the distribution of the random variable V_δ is the same as the posterior distribution of U with respect to the σ -algebra \mathcal{M}_δ .

Let ν_δ be the posterior distribution of U with respect to the σ -algebra \mathcal{M}_δ . Equivalently ν_δ is the distribution of V_δ in the Sobolev space $H^{\zeta_1}(N)$ where $\zeta_1 \leq \tau$. Let μ_δ be the distribution of the random variable W_δ which is independent of $U_\delta = T_\delta(M_\delta)$. Then the conditional expectation of the indicator function $\chi_{B_{\zeta_1}(U_\delta, R(\delta))}(U(\omega))$ with respect to \mathcal{M}_δ is

$$\begin{aligned} \mathbb{E}(\chi_{B_{\zeta_1}(U_\delta, R(\delta))} \mid \mathcal{M}_\delta)(\omega) &= \mathbb{P}(\{U \in B_{\zeta_1}(U_\delta, R(\delta)), R(\delta)\} \mid \mathcal{M}_\delta)(\omega) \\ &= \mathbb{P}(\{W_\delta \in B_{\zeta_1}(0, R(\delta))\})(\omega) \\ &= \mu_\delta(B_{\zeta_1}(0, R(\delta)))(\omega). \end{aligned} \tag{5.11}$$

Above $B_{\zeta_1}(0, R(\delta))$ denotes a ball in H^{ζ_1} of radius $R(\delta)$.

Let \mathcal{U} be the σ -algebra generated by U . By [8, theorem 10.2.2] there are regular conditional probabilities $\mathbf{P}(K \mid \mathcal{M}_\delta)(\omega)$ for all $K \in \mathcal{B}(H^{\zeta_1})$ and $\omega \in \Omega$ such that

$$\mathbf{P}(K \mid \mathcal{M}_\delta)(\omega) = \mathbb{E}(\chi_K(U) \mid \mathcal{M}_\delta)(\omega) \quad \text{a.s.}$$

Moreover, by applying [8, theorem 10.2.1] to the joint distribution of (U, M_δ) we see that there are such functions

$$(m, K) \mapsto \mathbb{P}_m(K) =: \mathbb{P}(\{U \in K\} \mid M_\delta = m),$$

defined for $m \in H^{-s}$ and $K \in \mathcal{B}(H^{\zeta_1})$, that

$$\mathbb{P}_{M_\delta(\omega)}(K) = \mathbf{P}(K \mid \mathcal{M}_\delta)(\omega) \quad \text{a.s.} \quad (5.12)$$

Using (5.11) and (5.12) we see that

$$\mathbb{P}(\{U \in B_{\zeta_1}(T_\delta(m), R(\delta))\} \mid M_\delta = m) = \mu_\delta(B_{\zeta_1}(0, R(\delta))).$$

Note that the right-hand side is in fact independent of m and depends only on δ . Next we will give a theorem for the credible sets

$$\mathbb{P}(\{V_\delta \in B_{\zeta_1}(U_\delta, R(\delta))\}) = \mathbb{P}(\{U \in B_{\zeta_1}(T_\delta(m_\delta), R(\delta))\} \mid M_\delta = m_\delta).$$

Theorem 4. Let U, \mathcal{E}, M_δ and A be defined as in theorem 1 and assume that $A^* : \mathcal{D}'(N) \rightarrow \mathcal{D}'(N)$ is invertible.

Let $\mathcal{M}_\delta = \sigma(M_\delta)$ and $\mathcal{U} = \sigma(U)$ be the σ -algebras generated by the random variable M_δ and U respectively. Then the posterior distribution of the random variable U with respect to the σ -algebra \mathcal{M}_δ can be given in terms of function

$$(m, K) \mapsto \mathbb{P}(\{U \in K\} \mid M_\delta = m),$$

where $m \in H^{-s}(N)$ and $K \in \mathcal{B}(H^{\zeta_1})$, see (5.12).

Take $\zeta_1 < \tau + t - t_0$ and $\alpha < \gamma/2$. Then if $R(\delta) = C_1 \delta^\alpha$ we have the following contraction:

$$\mathbb{P}(\{U \in B_{\zeta_1}(T_\delta(m_\delta), R(\delta))\} \mid M_\delta = m_\delta) \geq 1 - C\delta^{\gamma-2\alpha} \rightarrow 1$$

when $\delta \rightarrow 0$. The speed of contraction depends on ζ_1 :

- (i) If $\zeta_1 \leq -s - t_0$ then $\gamma = 2(t + r)/(t_0 + r)$.
- (ii) If $-s - t_0 \leq \zeta_1 < \tau + t - t_0$ then $\gamma = 2(\tau + t - t_0 - \zeta_1)/(t_0 + r)$.

Proof. We use below $R = R(\delta) = C_1 \delta^\alpha$ with some $\alpha > 0$ and denote

$$p_\delta = 1 - \mu_\delta(B_{\zeta_1}(0, R(\delta))).$$

To study what happens to p_δ we first notice that

$$\begin{aligned} \mathbb{E}(\|W_\delta\|_{H^{\zeta_1}}^2) &= \int_{H^{\zeta_1}} \|w\|_{H^{\zeta_1}}^2 d\mu_\delta(w) \\ &\geq \int_{\|w\|_{H^{\zeta_1}}^2 > (R(\delta))^2} \|w\|_{H^{\zeta_1}}^2 d\mu_\delta(w) \\ &\geq R(\delta)^2 \int_{\|w\|_{H^{\zeta_1}}^2 > (R(\delta))^2} d\mu_\delta(w) \\ &= (R(\delta))^2 p_\delta. \end{aligned} \quad (5.13)$$

Next we will prove that

$$\mathbb{E}(\|W_\delta\|_{H^{\zeta_1}}^2) = \text{Tr}_{H^{\zeta_1} \rightarrow H^{\zeta_1}} B_{W_\delta} \leq C\delta^\gamma$$

with some γ that depends on ζ_1 . Above we use the definition $B_{W_\delta} = C_{W_\delta}(I - \Delta)^{\zeta_1}$ where $C_{W_\delta} : H^{-\zeta_1} \rightarrow H^{\zeta_1}$ and

$$\langle C_{W_\delta} \phi, \psi \rangle_{H^{\zeta_1} \times H^{-\zeta_1}} = \mathbb{E}(\langle W_\delta, \phi \rangle_{H^{\zeta_1} \times H^{-\zeta_1}} \langle W_\delta, \psi \rangle_{H^{\zeta_1} \times H^{-\zeta_1}}).$$

As noted before when A^* is invertible we can write

$$C_{W_\delta} = F_\lambda^{-1} (A^* A)^{-1}, \quad (5.14)$$

where $F_\lambda = (A^* A)^{-1} C_U^{-1} + \lambda I$ and $\lambda = \delta^{-2}$.

We want to estimate

$$\mathbb{E}(\|W_\delta\|_{H^{\zeta_1}}^2) = \mathbb{E}(\|(I - \Delta)^{\zeta_1/2} W_\delta\|_{L^2}^2).$$

Let us define

$$W_{\delta, \zeta_1} = (I - \Delta)^{\zeta_1/2} W_\delta : \Omega \rightarrow L^2(N).$$

We can write the covariance operator of W_{δ, ζ_1}

$$C_{W_{\delta, \zeta_1}} = (I - \Delta)^{\zeta_1/2} C_{W_\delta} (I - \Delta)^{\zeta_1/2} : L^2 \rightarrow L^2.$$

Note that in L^2 we have $B_{W_{\delta, \zeta_1}} = C_{W_{\delta, \zeta_1}}$. Now we get for the trace

$$\begin{aligned} \text{Tr}_{L^2 \rightarrow L^2}((I - \Delta)^{\zeta_1/2} C_{W_\delta} (I - \Delta)^{\zeta_1/2}) &= \text{Tr}_{L^2 \rightarrow L^2}((I - \Delta)^{\zeta_1} C_{W_\delta}) \\ &= \text{Tr}_{L^2 \rightarrow L^2}((I - \Delta)^{-s} (I - \Delta)^{\zeta_1+s} C_{W_\delta}) \\ &\leq \text{Tr}_{L^2 \rightarrow L^2}((I - \Delta)^{-s}) \|(I - \Delta)^{\zeta_1+s} C_{W_\delta}\|_{L^2 \rightarrow L^2}. \end{aligned}$$

Using (5.14) we get

$$\begin{aligned} \|(I - \Delta)^{\zeta_1+s} C_{W_\delta}\|_{L^2 \rightarrow L^2} &= \|(I - \Delta)^{\zeta_1+s} F_\lambda^{-1} (A^* A)^{-1}\|_{0,0} \\ &\leq \|(I - \Delta)^{\zeta_1+s}\|_{2(\zeta_1+s),0} \|F_\lambda^{-1}\|_{-2t_0, 2(\zeta_1+s)} \|(A^* A)^{-1}\|_{0, -2t_0} \\ &\leq C \|F_\lambda^{-1}\|_{-2t_0, 2(\zeta_1+s)}. \end{aligned}$$

Above $F_\lambda^{-1} = ((A^* A)^{-1} C_U^{-1} + \lambda)^{-1} \in \Psi_p^m$ where $m = -2(t+r)$ and $p = 2(t_0+r)$. We want to use the norm estimates (4.7) and (4.8) so we write

$$\|F_\lambda^{-1}\|_{-2t_0, 2(\zeta_1+s)} = \|F_\lambda^{-1}\|_{-2t_0, -2t_0-\ell},$$

where $\ell = -2(s + \zeta_1 + t_0)$. To use the norm estimates we need to assume $\ell \geq m$, that is, $\zeta_1 \leq r - s + t - t_0 = \tau + t - t_0$.

First we assume that $\ell \geq 0$ which is true when $\zeta_1 \leq -s - t_0$. Then

$$\|F_\lambda^{-1}\|_{-2t_0, -2t_0-\ell} \leq C(1 + \lambda^{\frac{1}{2(t_0+r)}})^{-2(t+r)}$$

and $\|F_\lambda^{-1}\|_{-2t_0, -2t_0-\ell} \rightarrow 0$ when $\lambda \rightarrow \infty$ with all $t, t_0 > 0$ and $r \geq 0$.

Next we assume $\ell \leq 0$. Then for $-s - t_0 \leq \zeta_1 \leq \tau + t - t_0$ we get

$$\|F_\lambda^{-1}\|_{-2t_0, -2t_0-\ell} \leq C(1 + \lambda^{\frac{1}{2(t_0+r)}})^{2(s+\zeta_1+t_0)-2(t+r)}.$$

Now $\|F_\lambda^{-1}\|_{-2t_0, -2t_0-\ell} \rightarrow 0$ when $\lambda \rightarrow \infty$ if $\zeta_1 < \tau + t - t_0$.

We have proven that

$$\mathbb{E}(\|W_\delta\|_{H^{\zeta_1}}^2) \leq C\delta^\gamma,$$

where $\gamma = 2(t+r)/(t_0+r)$ if $\zeta_1 \leq -s - t_0$ and $\gamma = 2(\tau + t - t_0 - \zeta_1)/(t_0+r)$ if $-s - t_0 \leq \zeta_1 < \tau + t - t_0$. Hence using the above estimate and (5.13) we see that

$$p_\delta \leq \frac{C\delta^\gamma}{(C_1\delta^\alpha)^2} = C\delta^{\gamma-2\alpha}.$$

Above we have to assume $\alpha < \gamma/2$ to have convergence $p_\delta \rightarrow 0$ when $\delta \rightarrow 0$.

Finally, since we have denoted $u_\delta = T_\delta(m_\delta)$, we can conclude that with above choices for $R(\delta)$, γ and α

$$\begin{aligned} \mathbb{P}(\{V_\delta \in B_{\zeta_1}(U_\delta, R(\delta))\}) &= \mathbb{P}(\{U \in B_{\zeta_1}(T_\delta(m_\delta), R(\delta))\} \mid M_\delta = m_\delta) \\ &\geq 1 - C\delta^{\gamma-2\alpha} \rightarrow 1 \end{aligned}$$

when $\delta \rightarrow 0$. □

5.3. Discussion

Above, we have considered in the frequentist setting the case when the solution u^\dagger is an element of $H^\tau(N)$ with $\tau > 0$ and studied in theorems 2 and 3 the convergence of the MAP estimators and the contraction of the posterior distribution in $L^2(N)$.

In the Bayesian setting we have examined the case when the solution is a realisation of the random variable U . In theorems 1 and 4 we have studied the convergence of the MAP estimators and the contraction of the posterior distribution in $H^\zeta(N)$ with various values of ζ .

In classical regularisation theory for linear inverse problems, one is usually interested in the convergence of the optimisers of the minimisation problem (2.5) to the true solution in the space $H^r(N)$, $r > \tau + d/2$, as the noise level δ goes to zero. This gives restrictions to the measurement noise that can be considered. Summarising, our above statistical considerations concern the case where unknown and the noise are significantly less smooth than in the standard setting of the regularisation theory. In the recent regularisation theory inverse problems where the direct map A and the regularisation term are nonlinear have been studied extensively. It is interesting to ask how our analysis on the contraction of the posterior distribution could be generalised for such nonlinear inverse problems that corresponds to non-Gaussian statistical problems.

Acknowledgments

We would like to thank Petteri Piiroinen for helpful discussions. This work was supported by the Finnish Centre of Excellence in Inverse Problems Research 2012-2017 (Academy of Finland CoE-project 284715). In addition, HK was supported by Emil Aaltonen Foundation and EQUIP, grant EP/K034154/1, ML was supported by Academy of Finland, grant 273979, and SS was supported by Academy of Finland, project 141094.

Appendix A. Some examples of hypoelliptic operators

A linear partial differential operator P is hypoelliptic if for every distribution u such that $P(u)$ is C^∞ smooth also u is C^∞ . Every elliptic operator with smooth coefficients is hypoelliptic. The heat operator

$$Pu(x, t) = \partial_t u - k\Delta_x u, \quad (x, t) \in \mathbb{R}^d \times \mathbb{R}$$

and Kolmogorov operator [19, 30]

$$Pu(x, y, t) = \partial_{xx} u + x\partial_y u - \partial_t u, \quad (x, y, t) \in \mathbb{R}^3 \tag{A.1}$$

are examples of operators that are hypoelliptic but not elliptic. General Kolmogorov type hypoelliptic diffusion operators are used e.g. in the theory of kinetic equations, statistical physics and mathematical finance [17, 39].

The fact that (A.1) is hypoelliptic follows from Hörmander's theorem on hypoelliptic PDEs. Let (X_0, X_1, \dots, X_p) be real C^∞ vector fields in the d dimensional manifold N . If X and Y are two vector fields we define the bracket of X and Y by

$$[X, Y]f = X(Yf) - Y(Xf).$$

Note that $[X, Y]$ is a new vector field.

Definition 6 (Hörmander condition). We say that the Hörmander condition is satisfied if the real C^∞ vector fields (X_0, X_1, \dots, X_p) , $p \leq n$, in the manifold N generate a Lie algebra of rank $n = \dim N$ at every point $x \in N$.

This means that the vector fields

$$X_j, [X_{j_1}, X_{j_2}], [X_{j_1}, [X_{j_2}, X_{j_3}]], \dots$$

span a space that has the same dimension n as the manifold N at every point $x \in N$. Now we can formulate Hörmander's classical theorem [20].

Theorem 5 (Hörmander's theorem). *The operator*

$$P = \sum_{j=1}^p X_j^2 + X_0$$

defined on n dimensional manifold N is hypoelliptic if the vector fields (X_0, X_1, \dots, X_p) satisfy the Hörmander condition.

By writing

$$X_1 = \partial_x, \quad X_0 = x\partial_y - \partial_t, \quad [X_1, X_0] = \partial_y$$

we see that (A.1) is indeed hypoelliptic in \mathbb{R}^3 . Next we will give another, important example of vector fields satisfying Hörmander's condition

Example 4. Let us study Heisenberg group \mathbb{H} . Let Γ be a discrete subgroup of Heisenberg group \mathbb{H} such that \mathbb{H}/Γ is compact, see e.g. [12]. The orthonormal frame on \mathbb{R}^3 is given by the Lie vector fields

$$\begin{aligned} X &= \partial_x + \frac{1}{2}y\partial_t, \\ Y &= \partial_y - \frac{1}{2}x\partial_t, \\ Z &= \partial_t. \end{aligned}$$

We can easily see that

$$[X, Y] = Z.$$

Using theorem 5 we get that the sub-Laplacian

$$P = \frac{1}{2}(X^2 + Y^2)$$

on \mathbb{H}/Γ is hypoelliptic.

Example 5. One example of hypoelliptic inverse problem is the heat equation on a compact manifold $N = M \times \mathbb{R}$ where M is a closed two-dimensional manifold. Note that in this paper

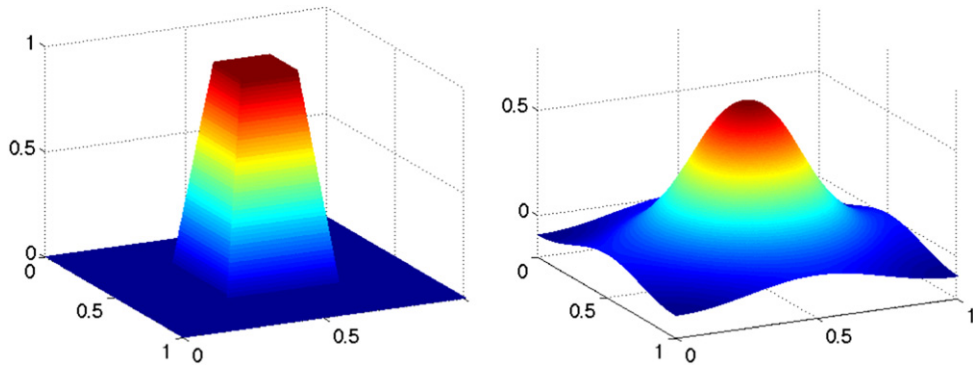


Figure B1. On the left the original piecewise linear function u^\dagger . On the right side the noiseless data $m^\dagger = Au^\dagger$.

we have considered the problem on a compact manifold, that is, our results are applicable in the case when the equation is periodic in time. We are interested in solving the heat sources $U(x, t)$ from the noisy measurements $M_\delta(x, t)$ of temperature $T(x, t)$, that is we want to solve U from

$$(\partial_t - \Delta_x)T(x, t) = U(x, t), \quad (\text{A.2})$$

$$M_\delta(x, t) = (\partial_t - \Delta_x)^{-1}U(x, t) + \delta\mathcal{E}. \quad (\text{A.3})$$

The operator $A = (\partial_t - \Delta_x)^{-1}$ is not elliptic but it is hypoelliptic of type (1, 2).

Such situations arise in non-invasive monitoring. Consider, for example, using a thermal camera to record video footage of a car with engine running. Let us model the metal surface of the car as a compact and closed two-dimensional manifold M . The running engine produces heat which we observe in the video data. The temperature on the car surface is modelled as the solution $T(x, t)$ defined on $N = M \times \mathbb{R}$. Equation (A.2) describes the conduction of heat along the car surface. The effect of the engine is simply modelled as the heat source term $U(x, t)$; recovering U will provide information about the state of the engine.

Appendix B. Computational example

Since the operator A does not have a continuous inverse operator $L^2 \rightarrow L^2$, the condition number of the matrix approximation \mathbf{A} of the operator A grows when the discretisation is refined. This is the very reason why regularisation is needed in the (numerical) solutions of the inverse problems.

Next we demonstrate the above results numerically and consider two-dimensional deblurring problem on \mathbb{T}^2 ,

$$M = AU + \delta\mathcal{E},$$

where $\mathcal{E}(\omega) \in H^{-s}$, $s > 1$ a.s. is normalised white noise and A is elliptic operator, smoothing of order 2

$$(Au)(x) = \mathcal{F}^{-1}((1 + |n|^2)^{-1}(\mathcal{F}u)(n))(x).$$

The true solution $u^\dagger \in H^1(\mathbb{T}^2)$, see section 2.1.2 for the frequentist interpretation, is a piecewise linear function presented in figure B1. We choose *a priori* distribution $N(0, C_U)$ where $C_U = (I - \Delta)^{-1}$. Then a random draw $U(\omega)$ from the prior distribution belongs to H^τ ,

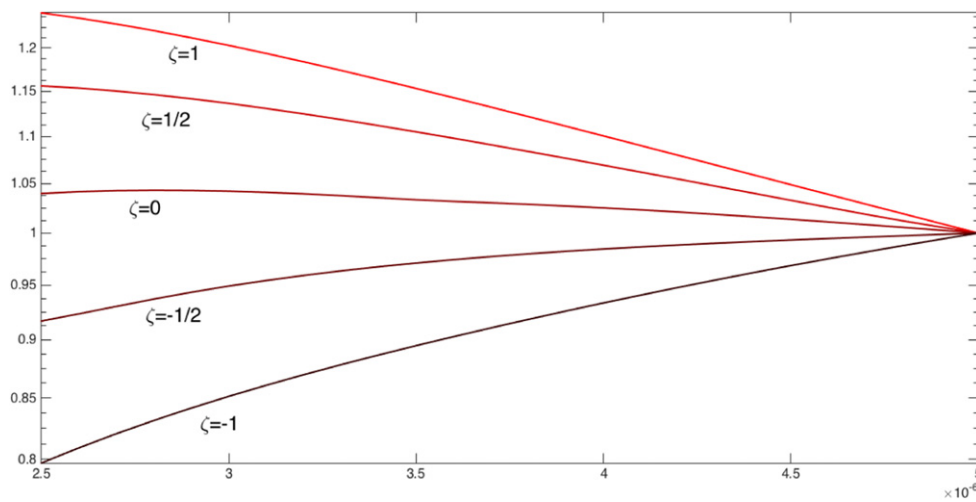


Figure B2. Normalised errors $c(\zeta)\|u^\dagger - u_\delta^\dagger\|_{H^\zeta(\mathbb{T}^2)}$ in logarithmic scale with different values of ζ . We use normalisation constants $c(\zeta) = 1/\|u^\dagger - u_{5 \cdot 10^{-6}}^\dagger\|_{H^\zeta}$. We observe that $u_\delta^\dagger \in H^1$ does not converge to $u^\dagger \in H^1$ in H^ζ when $\zeta > 0$.

where $\tau < 0$, with probability one. The Cameron–Martin space Z_U of a measurable mapping $U : \Omega \rightarrow X$ is defined by

$$Z_U = \{\phi \in X \mid \|\phi\|_{Z_U}^2 = \langle C_U^{-1}\phi, \phi \rangle_{X^* \times X} < \infty\}.$$

Cameron–Martin space can also be defined as

$$Z_U = \bigcap \{Y \mid Y \subset H^\tau \text{ linear subspace, } \mathbb{P}(\{U \in Y\}) = 1\}.$$

The approximated solutions U_δ belongs to Z_U and with the chosen *a priori* distribution we have $Z_U = H^1(\mathbb{T}^2)$.

Solving u from $Au(x) = m(x)$ corresponds to the solution of ordinary differential equation $(1 - \partial_x^2)m(x) = u(x)$ so A can be thought e.g. as a blurring operator.

The approximated solution to the problem is

$$u_\delta^\dagger = (A^*A + \delta^2(I - \Delta))^{-1}A^*m^\dagger.$$

We get from section 2.1.2 that

$$\lim_{\delta \rightarrow 0} \mathbb{E}_u \|u^\dagger - U_\delta^\dagger\|_{H^\zeta} = 0$$

when $\zeta < \tau < 0$. This behaviour can be seen even in numerical simulations when the discretisation is fine enough, see figure B2. In figure B3 we have compared the expected convergence rates given in formula (2.12) in theorem 1 to the computational convergence rates. In the numerical simulations in figures B2 and B3 we see that for the test case presented in figure B1 the convergence $u_\delta^\dagger \rightarrow u^\dagger$ in different Sobolev spaces follows well the mean convergence predicted by theorem 1.

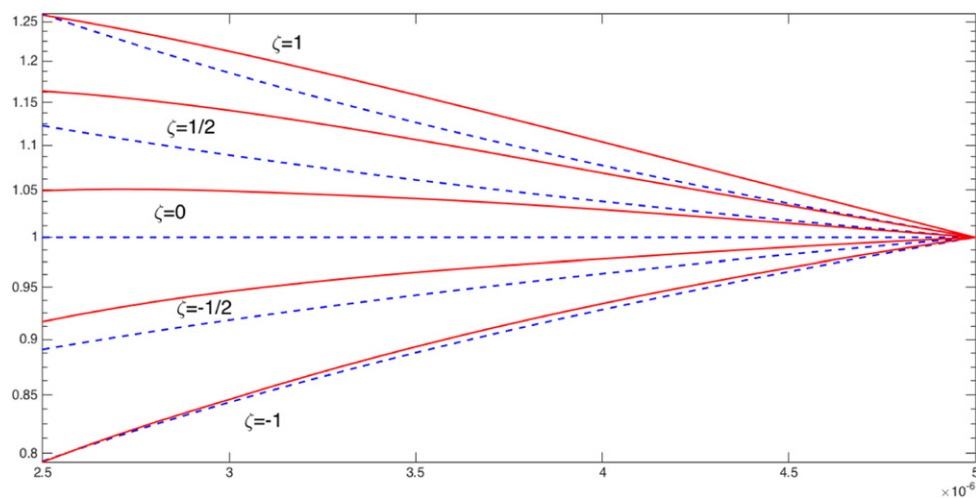


Figure B3. Normalised errors $c(\zeta)\|u^\dagger - u_\delta^\dagger\|_{H^\zeta(\mathbb{T}^2)}$ in logarithmic scale with different values of ζ . The normalised bounds (2.12) given in theorem 1 for the expectations $c_1(\zeta)\mathbb{E}\|U - U_\delta\|_{H^\zeta(\mathbb{T}^2)}$ are plotted with dashed lines. The normalised errors $c(\zeta)\|u^\dagger - u_\delta^\dagger\|_{H^\zeta(\mathbb{T}^2)}$, for the example u^\dagger given in figure B1, are plotted with solid lines.

References

- [1] Agapiou S, Larsson S and Stuart A M 2013 Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems *Stoch. Process. Appl.* **123** 3828–60
- [2] Agapiou S, Stuart A M and Zhang Y-X 2014 Bayesian posterior contraction rates for linear severely ill-posed inverse problems *J. Inverse Ill-Posed Problems* **22** 297–321
- [3] Bogachev V 1998 *Gaussian Measures (Mathematical Surveys and Monographs vol 62)* (Providence, RI: American Mathematical Society)
- [4] Castillo I and Nickl R 2013 Nonparametric Bernstein–von Mises theorems in Gaussian white noise *Ann. Stat.* **41** 1999–2028
- [5] Castillo I and Nickl R 2014 On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures *Ann. Stat.* **42** 1941–69
- [6] Cavalier L 2008 Nonparametric statistical inverse problems *Inverse Problems* **24** 034004
- [7] Dashti M, Law K J, Stuart A M and Voss J 2013 MAP estimators and their consistency in Bayesian nonparametric inverse problems *Inverse Problems* **29** 095017
- [8] Dudley R M 1989 *Real Analysis and Probability (Mathematics Series)* (Pacific Grove, CA: The Wadsworth & Brooks/Cole)
- [9] Engl H, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Netherlands: Springer)
- [10] Fitzpatrick B 1991 Bayesian analysis in inverse problems *Inverse Problems* **7** 675–702
- [11] Florens J-P and Simoni A 2016 Regularizing priors for linear inverse problems *Econometric Theory* **32** 71–121
- [12] Folland G 2004 Compact Heisenberg manifolds as CR manifolds *J. Geom. Anal.* **14** 521–32
- [13] Ghosal S 1997 A review of consistency and convergence of posterior distribution *Varanashi Symp. Bayesian Inference* (Banaras Hindu University)
- [14] Ghosal S, Ghosh J K and Van Der Vaart A W 2000 Convergence rates of posterior distributions *Ann. Stat.* **28** 500–31

- [15] Giné E and Nickl R 2015 *Mathematical Foundations of Infinite-Dimensional Statistical Models* (Cambridge: Cambridge University Press)
- [16] Groetsch C 1984 *The Theory of Tikhonov Regularization for Fredholm equations of the First Kind* (London: Pitman)
- [17] Helffer B and Nier F 2005 *Hypoelliptic Estimates and Spectral Theory for Fokker–Planck Operators and Witten Laplacians* (Berlin: Springer)
- [18] Hoffmann M, Rousseau J and Schmidt-Hieber J 2015 On adaptive posterior concentration rates *Ann. Stat.* **43** 2259–95
- [19] Hörmander L 1967 Hypoelliptic second order differential equations *Acta Math.* **119** 147–71
- [20] Hörmander L 1994 *The Analysis of Linear Partial Differential Operators III: Pseudo-Differential Operators* (Berlin: Springer)
- [21] Huang T-M 2004 Convergence rates for posterior distributions and adaptive estimation *Ann. Stat.* **32** 1556–93
- [22] Kaipio J and Somersalo E 2004 *Statistical and Computational Inverse Problems (Appl. Math. Sci. vol 160)* (Berlin: Springer)
- [23] Kekkonen H, Lassas M and Siltanen S 2014 Analysis of regularized inversion of data corrupted by white Gaussian noise *Inverse Problems* **30** 045009
- [24] Kirsch A 1996 *An Introduction to the Mathematical Theory of Inverse Problems* (New York: Springer)
- [25] Knapik B and Salomond J-B 2014 A general approach to posterior contraction in nonparametric inverse problems arXiv:1407.0335
- [26] Knapik B, Szabó B, van der Vaart A and van Zanten J 2012 Bayes procedures for adaptive inference in inverse problems for the white noise model *Probab. Theory Related Fields*
- [27] Knapik B, van der Vaart A and Van Zanten J 2011 Bayesian inverse problems with Gaussian priors *Ann. Stat.* **39** 2626–57
- [28] Knapik B, Van der Vaart A and Van Zanten J 2013 Bayesian recovery of the initial condition for the heat equation *Commun. Stat. Theory Methods* **42** 1294–313
- [29] Kolehmainen V, Lassas M, Niinimäki K and Siltanen S 2012 Sparsity-promoting Bayesian inversion *Inverse Problems* **28** 025005
- [30] Kolmogoroff A 1934 Zufällige bewegungen (zur theorie der brownischen bewegung) *Ann. Math.* **116**–7
- [31] Lasanen S 2002 Discretizations of generalized random variables with applications to inverse problems *PhD Thesis Ann. Acad. Sci. Fenn. Math. Diss.* University of Oulu, Oulu
- [32] Lasanen S 2012 Non-Gaussian statistical inverse problems: I. Posterior distributions *Inverse Problems Imaging* **6** 215–66
- [33] Lassas M, Saksman E and Siltanen S 2009 Discretization-invariant Bayesian inversion and Besov space priors *Inverse Problems Imaging* **3** 87–122
- [34] Lassas M and Siltanen S 2004 Can one use total variation prior for edge-preserving Bayesian inversion? *Inverse Problems* **20** 1537–64
- [35] Leahu H 2011 On the Bernstein–von Mises phenomenon in the Gaussian white noise model *Electron. J. Stat.* **5** 373–404
- [36] Lehtinen M, Päivärinta L and Somersalo E 1989 Linear inverse problems for generalised random variables *Inverse Problems* **5** 599–612
- [37] Mandelbaum A 1984 Linear estimators and measurable linear transformations on a Hilbert space *Z. Wahrscheinlichkeitstheor. Verwandte Geb.* **65** 385–97
- [38] Morozov V A, Nashed Z and Aries A 1984 *Methods for Solving Incorrectly Posed Problems* (Berlin: Springer)
- [39] Pascucci A 2011 *PDE and Martingale Methods in Option Pricing* vol 2 (Berlin: Springer)
- [40] Phillips D L 1962 A technique for the numerical solution of certain integral equations of the first kind *J. ACM (JACM)* **9** 84–97
- [41] Piironen P 2005 Statistical measurements, experiments and applications *PhD Thesis Ann. Acad. Sci. Fenn. Math. Diss.* University of Helsinki, Helsinki
- [42] Ray K 2013 Bayesian inverse problems with non-conjugate priors *Electron. J. Stat.* **7** 2516–49
- [43] Ray K 2014 Adaptive Bernstein–von Mises theorems in Gaussian white noise arXiv:1407.3397
- [44] Rozanov I A 1971 *Infinite-Dimensional Gaussian Distributions* (Providence, RI: American Mathematical Society)
- [45] Rozanov Y 1982 *Markov Random Fields* (Berlin: Springer)
- [46] Rudin W 1987 *Real and Complex Analysis* 3rd edn (New York: McGraw-Hill)

- [47] Schuster T, Kaltenbacher B, Hofmann B and Kazimierski K S 2012 *Regularization Methods in Banach Spaces* vol 10 (Berlin: de Gruyter & Co)
- [48] Shen X and Wasserman L 2001 Rates of convergence of posterior distributions *Ann. Stat.* **29** 687–714
- [49] Shubin M A and Andersson S I 1987 *Pseudodifferential Operators and Spectral Theory* vol 200 (Berlin: Springer)
- [50] Stuart A M 2010 Inverse problems: a Bayesian perspective *Acta Numer.* **19** 451–559
- [51] Szabo B, van der Vaart A and van Zanten H 2015 Honest Bayesian confidence sets for the l2-norm *J. Stat. Plan. Inference* **166** 36–51
- [52] Szabó B *et al* 2015 Frequentist coverage of adaptive nonparametric Bayesian credible sets *Ann. Stat.* **43** 1391–428
- [53] Tikhonov A N 1943 On the stability of inverse problems *Dokl. Akad. Nauk SSSR* **39** 176–9
- [54] Tikhonov A N 1963 Solution of incorrectly formulated problems and the regularization method *Sov. Mathe. Dokl.* **4** 261–5
- [55] Tikhonov A N 1995 *Numerical Methods for the Solution of Ill-Posed Problems* vol 328 (Berlin: Springer)
- [56] van der Vaart A W 2000 *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*
- [57] van der Vaart A W and van Zanten J H 2008 Rates of contraction of posterior distributions based on Gaussian process priors *Ann. Stat.* **36** 1435–63
- [58] Vollmer S J 2013 Posterior consistency for Bayesian inverse problems through stability and regression results *Inverse Problems* **29** 125011