

Analysis of regularized inversion of data corrupted by white Gaussian noise

Hanne Kekkonen, Matti Lassas and Samuli Siltanen

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

E-mail: hanne.kekkonen@helsinki.fi

Received 27 November 2013, revised 6 February 2014

Accepted for publication 11 February 2014

Published 5 March 2014

Abstract

Tikhonov regularization is studied in the case of linear pseudodifferential operator as the forward map and additive white Gaussian noise as the measurement error. The measurement model for an unknown function $u(x)$ is

$$m(x) = Au(x) + \delta\varepsilon(x),$$

where $\delta > 0$ is the noise magnitude. If ε was an L^2 -function, Tikhonov regularization gives an estimate

$$T_\alpha(m) = \arg \min_{u \in H^r} \{ \|Au - m\|_{L^2}^2 + \alpha \|u\|_{H^r}^2 \}$$

for u where $\alpha = \alpha(\delta)$ is the regularization parameter. Here penalization of the Sobolev norm $\|u\|_{H^r}$ covers the cases of standard Tikhonov regularization ($r = 0$) and first derivative penalty ($r = 1$). Realizations of white Gaussian noise are almost never in L^2 , but do belong to H^s with probability one if $s < 0$ is small enough. A modification of Tikhonov regularization theory is presented, covering the case of white Gaussian measurement noise. Furthermore, the convergence of regularized reconstructions to the correct solution as $\delta \rightarrow 0$ is proven in appropriate function spaces using microlocal analysis. The convergence of the related finite-dimensional problems to the infinite-dimensional problem is also analysed.

Keywords: regularization, inverse problem, white noise, pseudodifferential operator

1. Introduction

1.1. Discrete and continuous regularization

Consider the following continuous model for indirect measurements:

$$m = Au + \text{noise}, \tag{1.1}$$

where the data m and the quantity of interest u are real-valued functions of d real variables and A is a bounded linear operator. A large class of practical measurements can be modelled by operators A arising from partial differential equations of mathematical physics. We focus on ill-posed inverse problems where A does not have a continuous inverse.

Physical measurement devices produce a discrete data vector $\mathbf{m} \in \mathbb{R}^k$, which we model by adding a linear operator P_k to (1.1):

$$\mathbf{m} := P_k(Au) + P_k(\text{noise}). \quad (1.2)$$

Furthermore, practical solution of the inverse problem calls for a discrete representation of the unknown u . This can be done using some computationally feasible approximation of the form $\mathbf{u} = T_n u \in \mathbb{R}^n$, for example Fourier series truncated to n terms. The practical inverse problem is now

$$\text{given } \mathbf{m}, \text{ compute a noise-robust approximation to } \mathbf{u}. \quad (1.3)$$

We study the most common computational approach to (1.3), namely classical Tikhonov regularization defined by

$$T_\alpha(\mathbf{m}) := \arg \min_{\mathbf{u} \in \mathbb{R}^n} \{ \|\mathbf{A}\mathbf{u} - \mathbf{m}\|_2^2 + \alpha \|\mathbf{L}\mathbf{u}\|_2^2 \}. \quad (1.4)$$

Here $\mathbf{A} = P_k A T_n$ is a $k \times n$ matrix approximation to the operator A , and $0 < \alpha < \infty$ is the *regularization parameter*. The matrix \mathbf{L} is used to introduce *a priori* information to the inversion. For example,

- (a) $\mathbf{L} = \mathbf{I}$, the identity matrix, models the *a priori* information that u is not very large in norm,
- (b) $\mathbf{L} = \mathbf{I} + \mathbf{D}$, where \mathbf{D} is a finite-difference first-order derivative matrix, models the *a priori* information that u is continuously differentiable and u or its derivative are not very large in square norm.

Our aim is to provide new analytic insight to the relationship between the continuous model (1.1) and practical inversion based on (1.4) in the case of Gaussian noise.

Note that the reconstruction $T_\alpha(\mathbf{m})$ given by (1.4) depends on both k and n . Practical computational inversion may involve modifying both of them: updating the measurement device changes the number k of data points, and refining the computational grid in the hope of extra accuracy will increase n . Furthermore, sometimes the most efficient numerical algorithm is based on a multigrid strategy, involving the computation of $T_\alpha(\mathbf{m})$ with several different n .

Since there is a common continuous inverse problem behind the discrete model, it is desirable that the reconstruction $T_\alpha(\mathbf{m})$ converges to a meaningful limit as $k, n \rightarrow \infty$. Such convergence would also ensure that the dependency of $T_\alpha(\mathbf{m})$ on k and n is stable, at least for large enough values. Therefore, we discuss a continuous version of (1.4) based directly on the ideal model (1.1).

Under certain assumptions (including that m should be an L^2 -function) the finite-dimensional problem (1.4) Γ -converges as $n, k \rightarrow \infty$ to the following infinite-dimensional minimization problem in a Sobolev space H^r :

$$\arg \min_{u \in H^r} \{ \|m - Au\|_{L^2}^2 + \alpha \|u\|_{H^r}^2 \}. \quad (1.5)$$

See section 5 below for a proof. In (1.5) the case $r = 0$ corresponds to (a) and $r = 1$ corresponds, roughly, to (b) above. However, formula (1.5) only makes sense if the noise in (1.1) is square-integrable. This brings us to the main topic of the paper: noise modelling.

1.2. Properties of white noise

Next we will give the definitions for the discrete and continuous white noise and describe the ‘white noise paradox’ arising from the infinite L^2 -norm of the natural limit of white Gaussian noise in \mathbb{R}^k when $k \rightarrow \infty$.

We model the k -dimensional noise in (1.2) as $P_k(\text{noise}) = \delta \mathbf{e}$, where $\delta > 0$ plays the role of noise amplitude. The vector $\mathbf{e} \in \mathbb{R}^k$ is a realization of a \mathbb{R}^k -valued Gaussian random variable $\mathbf{E} = \mathbf{E}^{(k)}$ having mean zero and unit variance: $\mathbf{E}^{(k)} \sim N(0, I)$. In terms of a probability density function we have

$$\pi_{\mathbf{E}^{(k)}}(E) = c \exp\left(-\frac{1}{2}\|E\|_2^2\right), \quad E \in \mathbb{R}^k, \quad \|E\|_2 = \left(\sum_{j=1}^k E_j^2\right)^{1/2}. \quad (1.6)$$

The appearance of $\|\cdot\|_2$ in (1.6) is the reason why square norm is used in the data fidelity term $\|\mathbf{A}\mathbf{u} - \mathbf{m}\|_2^2$ of (1.4). The above noise model is appropriate for example for photon counting under high radiation intensity, see e.g. [24, 43].

Let us relate the above to the continuous model (1.1). We take $u(x)$ and $m(x)$ to be functions defined on a closed, compact d -dimensional manifold N , and the operator A to be a pseudodifferential operator (Ψ DO). Furthermore, the noise in (1.1) is modelled as $\delta\varepsilon(x)$, where $\delta > 0$ is the noise amplitude and $\varepsilon = \varepsilon(x)$ is a realization of normalized Gaussian white noise $W(x)$.

Rigorous treatment of white noise on N is based on generalized functions (distributions). We denote the pairing of a distribution $f \in \mathcal{D}'(N)$ and a test function $\phi \in C^\infty(N)$ by $\langle f, \phi \rangle$. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. A random generalized function $V = V(x, \omega)$, where $x \in N$ and $\omega \in \Omega$, on N is a measurable map $V : \Omega \rightarrow \mathcal{D}'(N)$. Below, following the tradition used in study of stochastic processes, we often omit the ω variable and just denote a random generalized function by $V(x)$.

White noise $W(x)$ is a random generalized function on N such that the inner products $\langle W, \phi \rangle$ are Gaussian random variables for all $\phi \in C^\infty(N)$, $\mathbb{E}W = 0$, and

$$\mathbb{E}(\langle W, \phi \rangle \langle W, \psi \rangle) = \langle \phi, C_W \psi \rangle_{L^2(N)} \quad \text{for } \phi, \psi \in C^\infty(N). \quad (1.7)$$

The covariance operator C_W of Gaussian white noise is the identity operator. Then W can be considered as a function $W : \Omega \rightarrow \mathcal{D}'(N)$ where Ω is the probability space. A realization of W is the generalized function $x \mapsto W(x, \omega)$ on N with a fixed $\omega \in \Omega$.

Below, we consider the case when

$$P_k(f) = (\langle f, \phi_j \rangle)_{j=1}^k, \quad (1.8)$$

where $\phi_j \in C^\infty(N)$ are such that $(\phi_j)_{j=1}^\infty$ is an orthogonal basis in $L^2(N)$. Then $P_k(\delta\varepsilon(x)) = \delta \mathbf{e} \in \mathbb{R}^k$ with \mathbf{e} as above. For example, when N is a d -dimensional torus, P_k can be the truncation of the Fourier series. See section 5 for a detailed discussion on discrete and continuous noise models.

Now we can state the main motivation behind this study. The probability density function of W is often *formally* written in the form

$$\pi_W(w) = c \exp\left(-\|w\|_{L^2(N)}^2/2\right). \quad (1.9)$$

However, despite formula (1.9), the realizations of the white Gaussian noise are almost surely not in $L^2(N)$. Thus we cannot use formula (1.5) when the error in the measurement m is white Gaussian noise. Let us illustrate this ‘white noise paradox’ by a simple example.

Example 1. Let W be normalized Gaussian white noise defined on the d -dimensional torus $\mathbb{T}^d = (\mathbb{R}/(2\pi\mathbb{Z}))^d$. The Fourier coefficients of W are normally distributed with variance one,

that is, $\langle W, e_{\vec{\ell}} \rangle \sim N(0, 1)$, where $e_{\vec{\ell}}(x) = e^{i\vec{\ell} \cdot x}$ and $\vec{\ell} \in \mathbb{Z}^d$. Hence

$$\|W\|_{L^2(\mathbb{T}^d)}^2 = \sum_{\vec{\ell} \in \mathbb{Z}^d} |\langle W, e_{\vec{\ell}} \rangle|^2 < \infty \quad \text{with probability zero.}$$

This implies that $W \in L^2(\mathbb{T}^d)$ with probability zero. However, when $s < -d/2$ we have $W \in H^s(\mathbb{T}^d)$ almost surely (that is, with probability one). This can be seen as follows: the Sobolev norm of W is given by

$$\|W\|_{H^s(\mathbb{T}^d)}^2 = \sum_{\vec{k} \in \mathbb{Z}^d} (1 + |\vec{k}|^2)^s |\langle W, e_{\vec{k}} \rangle|^2. \quad (1.10)$$

By [20, lemma 4.16], the infinite sum in (1.10) converges almost surely if and only if

$$\sum_{\vec{\ell} \in \mathbb{Z}^d} (1 + |\vec{\ell}|^2)^{2s} \mathbb{E} |\langle W, e_{\vec{\ell}} \rangle|^4 \leq C \sum_{\vec{\ell} \in \mathbb{Z}^d} (1 + |\vec{\ell}|^2)^{2s} < \infty.$$

This shows that the realizations of white noise W are almost surely in the space $H^s(\mathbb{T}^d)$ if and only if $s < -d/2$. In particular for $s \geq -d/2$ the function $x \mapsto W(x, \omega)$ is in $H^s(\mathbb{T}^d)$ only when $\omega \in \Omega_0 \subset \Omega$ where $\mathbb{P}(\Omega_0) = 0$.

Even though the previous example is proven in \mathbb{T}^d we note that the same result is valid in all open bounded subsets $D \subset \mathbb{R}^d$.

1.3. Main result

Let us again consider a general closed d -dimensional Riemannian manifold (N, g) and let $\Delta = \Delta_g$ be the Laplace operator on N . Furthermore, let A be a pseudodifferential operator. Consider the following measurement model:

$$m = Au + \delta\varepsilon, \quad (1.11)$$

where $\varepsilon \in H^s(N)$ with $s < -d/2$ is a realization of white noise.

The pseudodifferential operator A can be, for example,

$$Au(x) = \int_N \mathcal{A}(x, z)u(z) dz$$

where $\mathcal{A} \in C^\infty((N \times N) \setminus \text{diag}(N))$ and in an open neighbourhood $U \Subset N \times N$ of the $\text{diag}(N) = \{(x, x); x \in N\}$, we have

$$\mathcal{A}(x, z) = \frac{b(x, z)}{d_g(x, z)^p}, \quad (x, z) \in U$$

where d_g is a distance function, $p < d$, $b \in C^\infty(U)$ and $b(x, x) \neq 0$. In this case \mathcal{A} is a pseudodifferential operator of order $-d + p < 0$.

Let us now modify formula (1.5) to arrive at something useful for white Gaussian noise. Expand the data fidelity term like this: $\|m - Au\|_{L^2(N)}^2 = \|Au\|_{L^2(N)}^2 - 2\langle m, Au \rangle + \|m\|_{L^2(N)}^2$. Simply omitting the ‘constant term’ $\|m\|_{L^2(N)}^2$ leads to the definition

$$T_\alpha(m) := \arg \min_{u \in H^r(N)} \{ \|Au\|_{L^2(N)}^2 - 2\langle m, Au \rangle + \alpha \|u\|_{H^r(N)}^2 \}, \quad (1.12)$$

where we can interpret $\langle m, Au \rangle$ as a suitable duality pairing instead of $L^2(N)$ inner product. When A is a pseudodifferential operator of order $-t < s + r$, we can define $\langle m, Au \rangle = \langle m, Au \rangle_{H^s(N) \times H^{-s}(N)}$.

It is well-known that the solution of the finite-dimensional problem (1.4) can be calculated using the following formula:

$$T_\alpha(\mathbf{m}) = (\mathbf{A}^T \mathbf{A} + \alpha \mathbf{L}^T \mathbf{L})^{-1} \mathbf{A}^T \mathbf{m}. \quad (1.13)$$

The regularized solution of the continuous problem (1.12) is

$$T_\alpha(m) = (A^*A + \alpha(I - \Delta)^r)^{-1}A^*m. \quad (1.14)$$

The regularization parameter is chosen to be a function of the noise amplitude: $\alpha(\delta) = \alpha_0\delta^\kappa$, where $\alpha_0 > 0$ is a constant and $\kappa > 1$. We will now formulate the main theorem of this paper, concerning the continuous regularized solution (1.14).

Theorem 1. *Let N be a d -dimensional closed manifold and $u \in H^r(N)$ with $r \geq 0$. Here $\|u\|_{H^r(N)} := \|(I - \Delta)^{r/2}u\|_{L^2(N)}$. Let $\varepsilon \in H^s(N)$ with some $s < -d/2$ and consider the measurement*

$$m_\delta = Au + \delta\varepsilon, \quad (1.15)$$

where $A \in \Psi^{-t}$, is an elliptic pseudodifferential operator of order $-t$ on the manifold N with $t > \max\{0, -s - r\}$ and $\delta \in \mathbb{R}_+$. Assume that $A : L^2(N) \rightarrow L^2(N)$ is injective. The regularization parameter is chosen to be $\alpha(\delta) = \alpha_0\delta^\kappa$, where $\alpha_0 > 0$ is a constant and $\kappa > 1$.

Take $s_1 \leq s - t + (t + r)/\kappa$. Then the following convergence takes place in $H^{s_1}(N)$ -norm:

$$\lim_{\delta \rightarrow 0} T_{\alpha(\delta)}(m_\delta) = u.$$

Furthermore, we have the following estimates for the speed of convergence.

(i) If $s_1 \leq s - t$ then

$$\|T_{\alpha(\delta)}(m_\delta) - u\|_{H^{s_1}} \leq C \max\{\delta^\nu, \delta^{1+\omega/2}\}.$$

(ii) If $s - t \leq s_1 < s - t + (t + r)/\kappa$ then

$$\|T_{\alpha(\delta)}(m_\delta) - u\|_{H^{s_1}} \leq C \max\{\delta^\nu, \delta^{1+\omega}\}.$$

Above, the exponents are $\nu = \frac{1}{2}\kappa(r - \zeta)/(t + r)$, where $\zeta = \max\{s_1, -r - 2t\}$, and $\omega = \kappa(s - t - s_1)/t + r$.

Notice that in case (i) $\omega \geq 0$ and in case (ii) $\omega \leq 0$. The different convergence speeds (i) and (ii) show the trade-off between 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. In section 4 we give two counterexamples to show that even though $u \in H^r$ and $T_{\alpha(\delta)}(m_\delta) \in H^r$ the regularized solution does not converge to the real solution in H^r -norm.

1.4. Literature review

There are two main ways in inverse problems literature for modelling noise. The first approach based on the deterministic regularization techniques is to assume that the noise is deterministic and small. In that case one has a norm estimate of the noise and can study what happens when $\|\text{noise}\|_{L^2} \rightarrow 0$. This approach was originated by Tikhonov [46, 47], and studied in depth in [5, 11, 14, 21, 37, 38, 48]. The second approach to handling the noise is based on statistical point of view. The statistical modelling of noise in the inverse problems started in the early papers of [12, 13, 44, 45] and it is notable that with this approach one needs not assume smallness of the noise. For some recent references of the frequentist view of statistical problems see [3, 16, 32, 36]. Another statistical way to study inverse problems with random noise is based on Bayesian approach where m , x and ε are considered to be realizations of random variables, see [6, 15, 19, 25–31, 39–41].

The deterministic regularization and statistical approaches differ both in assumptions and techniques. This paper aims to bridge the gap between them. Our results are closely related

to earlier studies of Eggermont, LaRiccia, and Nashed [8–10], who studied weakly bounded noise. They assume that the noise is a L^2 -function and discuss regularization techniques when the noise tends to zero in the weak topology of L^2 . This kind of relaxed assumption of noise covers small low frequency noise and large high frequency noise. However, even though $\delta\varepsilon$ tends to zero in weak sense as $\delta \rightarrow 0$ when ε is a realization of the normalized white noise, this type of noise lies outside the definition of the weakly bounded noise as ε is not almost surely L^2 -valued.

A related approach of smoothing the noise before the analysis is described in [34, 35]. A similar regularization method where no smoothness of the operator A is assumed, but instead the regularization method is modified, is studied in [7]. Another possible approach to deal with white noise is to first perform a data projection step and then proceed to Tikhonov regularization [22, 23]. Also, Hohage and Werner have earlier studied inverse problems taking into account the fact that white noise is not square-integrable in [17].

Our new results are different from all of those previous studies. Our approach aims to study the effect of the continuous white noise not being an L^2 -function in Tikhonov regularization (1.5) instead of modifying the problem by altering the regularization method or assumptions.

2. Analysis of the translation-invariant case

Before giving the general proof of theorem 1 in section 3 we motivate the proof by proving a similar kind of lemma for translation-invariant case.

The regularized solution we are studying is of the form

$$T_{\alpha(\delta)}(m) := \arg \min_{u \in H^r(\mathbb{T}^d)} \{ \|Au\|_{L^2(\mathbb{T}^d)}^2 - 2\langle m, Au \rangle + \alpha \|(I - \Delta)^{r/2} u\|_{L^2(\mathbb{T}^d)}^2 \},$$

where $\alpha(\delta) = \alpha_0 \delta^\kappa$, for some constant α_0 and $\kappa > 1$. As mentioned before solution to this is

$$T_{\alpha(\delta)}(m) = (A^*A + \alpha(I - \Delta)^r)^{-1} A^* m. \quad (2.1)$$

Let us consider the case when $\alpha = \alpha_0 \delta^2$, where α_0 is a constant, and A is an elliptic pseudodifferential operator of order $-t < 0$ that commutes with translations. Then, in $L^2(\mathbb{T}^d)$ we have that $B = A^*A \geq c_1(I - \Delta)^{-t}$. As A and B commute with translations they are Fourier multipliers,

$$\widehat{Au}(n) = a(n)\widehat{u}(n)$$

and since A is elliptic there is $n_0 > 0$ so that

$$c_1|n|^{-t} \leq |a(n)| \leq c_2|n|^{-t}, \quad \text{for } |n| > n_0.$$

The symbol z_δ of $Z = A^*A + \alpha_0\delta^2(I - \Delta)^r$ is

$$z_\delta(n) = |a(n)|^2 + \alpha_0\delta^2(1 + n^2)^r$$

and thus

$$z_\delta(n) \geq \max(|a(n)|^2, \alpha_0\delta^2(1 + n^2)^r).$$

If $0 < \beta < \frac{1}{2}$ and $|n| > n_0$

$$\begin{aligned} z_\delta(n) &\geq |a(n)|^{2(1-\beta)} (\alpha_0\delta^2(1 + n^2)^r)^\beta \\ &\geq c_3|n|^{-2(1-\beta)t+2r\beta} \delta^{2\beta} \alpha_0^\beta. \end{aligned}$$

Now when $s < -d/2$ we have

$$\varepsilon \in H^s(\mathbb{T}^d).$$

Thus writing

$$\begin{aligned} T_{\alpha(\delta)}(m_\delta) &= (A^*A + \alpha_0\delta^2(I - \Delta)^r)^{-1}(A^*(Au + \delta\varepsilon)) \\ &= (A^*A + \alpha_0\delta^2(I - \Delta)^r)^{-1}A^*Au + (A^*A + \alpha_0\delta^2(I - \Delta)^r)^{-1}A^*(\delta\varepsilon) \end{aligned} \quad (2.2)$$

we see that

$$T_{\alpha(\delta)}(m_\delta) = v_\delta + w_\delta$$

where

$$\begin{aligned} \widehat{v}_\delta(n) &= \frac{1}{z_\delta(n)} |a(n)|^2 \widehat{u}(n), \\ \widehat{w}_\delta(n) &= \frac{1}{z_\delta(n)} \overline{a(n)} \widehat{\varepsilon}(n) \delta. \end{aligned}$$

Here,

$$\left| \frac{1}{z_\delta(n)} |a(n)|^2 \right| \leq 1, \quad \lim_{\delta \rightarrow 0} \frac{1}{z_\delta(n)} |a(n)|^2 = 1$$

and thus if $u \in H^r(\mathbb{T}^d)$ by dominated convergence theorem

$$\lim_{\delta \rightarrow 0} v_\delta = u, \quad \text{in } H^r(\mathbb{T}^d).$$

Above the limit speed of convergence can be analysed using the standard regularization theory [11] and the fact that

$$\begin{aligned} v_\delta &= (A^*A + \alpha_0\delta^2(I - \Delta)^r)^{-1}A^*Au \\ &= u - \alpha_0\delta^2(A^*A + \alpha_0\delta^2(I - \Delta)^r)^{-1}(I - \Delta)^r u. \end{aligned}$$

We can use the fact that $Z = A^*A + \alpha_0\delta^2(I - \Delta)^r \geq \alpha_0\delta^2(I - \Delta)^r$ and write

$$\begin{aligned} \|Z^{-1/2}(I - \Delta)^r u\|_{L^2} &\leq (\alpha_0\delta^2)^{-1/2} \|(I - \Delta)^{r/2} u\|_{L^2} \\ &\leq (\alpha_0\delta^2)^{-1/2} \|u\|_{H^r}. \end{aligned}$$

We also have the inequality $Z \geq A^*A \geq c_1(I - \Delta)^{-t}$. When $r > 0$ we can define $\eta = t/(2r+2t)$ and $\gamma = r/(2r+2t)$ so that, $\gamma + \eta = 1/2$, $t\gamma - r\eta = 0$. We get

$$\begin{aligned} \|Z^{-1}(I - \Delta)^r u\|_{L^2} &= \|Z^{-\gamma-\eta-1/2}(I - \Delta)^r u\|_{L^2} \\ &\leq (\alpha_0\delta^2)^{-1/2} \|Z^{-\gamma-\eta}(I - \Delta)^{r/2} u\|_{L^2} \\ &\leq (\alpha_0\delta^2)^{-1/2} \|(c_1(I - \Delta)^{-t})^{-\gamma} (\alpha_0\delta^2(I - \Delta)^r)^{-\eta} (I - \Delta)^{r/2} u\|_{L^2} \\ &\leq c_1^{-\gamma} (\alpha_0\delta^2)^{-\eta-1/2} \|(I - \Delta)^{r/2} u\|_{L^2} \\ &\leq c_1^{-\gamma} (\alpha_0\delta^2)^{-\eta-1/2} \|u\|_{H^r}. \end{aligned}$$

Hence we obtain

$$\begin{aligned} \|\alpha_0\delta^2 Z^{-1}(I - \Delta)^r u\|_{L^2} &\leq c_1^{-\gamma} (\alpha_0\delta^2)^{1/2-\eta} \|u\|_{H^r} \\ &= c_1^{-\gamma} \delta^{\frac{t}{t+r}} \|u\|_{H^r}. \end{aligned}$$

On the other hand,

$$\begin{aligned} \left| \frac{1}{z_\delta(n)} \overline{a(n)} \delta \right| &\leq \frac{1}{c_3 |n|^{-2(1-\beta)t+2r\beta} \delta^{2\beta} \alpha_0^\beta} c_2 |n|^{-t} \delta \\ &\leq c_4 |n|^{(1-2\beta)t-2r\beta} \delta^{1-2\beta} \alpha_0^{-\beta}. \end{aligned}$$

Hence

$$\|w_\delta\|_{H^{s_1}(\mathbb{T}^d)} \leq c_5 \delta^{1-2\beta}$$

where $s_1 \leq s - (1 - 2\beta)t + 2r\beta$. Because we proved the convergence of v_δ in L^2 we have to have $s_1 \leq 0$. This is true at least when $s \leq -r$. Thus adding the above results together we can formulate the next lemma.

Lemma 2. Let $u \in H^r(\mathbb{T}^d)$, $r > 0$, be Gaussian distributed, $\varepsilon \in H^s(\mathbb{T}^d)$, $s < \max\{-d/2, -r\}$, and

$$m_\delta = Au + \delta\varepsilon$$

where $A : H^r(\mathbb{T}^d) \rightarrow H^{r+t}(\mathbb{T}^d)$, $t > \max\{0, -s-r\}$, is an elliptic pseudodifferential operator of order $-t < 0$ that commutes with translations. We assume that $\alpha(\delta) = \alpha_0\delta^2$. Then for the regularized solution

$$T_{\alpha(\delta)}(m) = (A^*A + \alpha(I - \Delta)^r)^{-1}A^*m$$

of u we have

$$\lim_{\delta \rightarrow 0} T_{\alpha(\delta)}(m_\delta) = u, \quad \text{in } H^{s_1}(\mathbb{T}^d)$$

where $s_1 \leq s - (1 - 2\beta)t + 2r\beta \leq 0$ and $0 < \beta < 1/2$. Furthermore we have the following estimate of the speed of convergence

$$\|T_{\alpha(\delta)}(m_\delta) - u\|_{H^{s_1}} \leq C \max\{\delta^{\frac{r}{t+r}}, \delta^{1-2\beta}\}.$$

Proof. The convergence is immediate consequence of the above results. For the convergence speed we get

$$\begin{aligned} \|T_{\alpha(\delta)}(m_\delta) - u\|_{H^{s_1}} &= \|-\alpha_0\delta^2 T^{-1}(I - \Delta)^r u + w_\delta\|_{H^{s_1}} \\ &\leq \|\alpha_0\delta^2 T^{-1}(I - \Delta)^r u\|_{L^2} + \|w_\delta\|_{H^{s_1}} \\ &\leq C_1(\alpha_0\delta^2)^{1/2-\eta} + C_2\delta^{1-2\beta}\alpha_0^{-\beta} \\ &\leq C_3 \max\{\delta^{\frac{r}{t+r}}, \delta^{1-2\beta}\} \end{aligned}$$

where $\eta = t/2(t+r)$. □

3. Proof of the main theorem

Here we study the general case where A is an elliptic pseudodifferential operator of order $-t < 0$. We denote $H^s(N) = H^s$ and $L^2(N) = L^2$ where N is a closed manifold and $\dim N = d$. As in the previous example we have

$$\begin{aligned} T_{\alpha(\delta)}(m_\delta) &= T^{-1}A^*Au + T^{-1}A^*(\delta\varepsilon) \\ &= u - \alpha T^{-1}(I - \Delta)^r u + T^{-1}A^*(\delta\varepsilon) \end{aligned} \quad (3.1)$$

where $T = A^*A + \alpha(I - \Delta)^r$.

First we will show that $B = A^*A$ is invertible. We define $A^* : L^2(N) \rightarrow L^2(N)$ as the adjoint of an operator $A : L^2(N) \rightarrow L^2(N)$. We assume that $A : L^2(N) \rightarrow L^2(N)$ is one-to-one. If $Bu = 0$ then

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

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

Next we recall the fact that an elliptic operator $B \in \Psi^{-2t}(N)$ is a Fredholm operator and $\text{index}(B) = 0$ ([18] theorem 19.2.1). Indeed index of a Fredholm operator B is

$$\text{index}(B) = \dim(\text{Ker } B) - \dim(\text{Coker } B). \quad (3.2)$$

If $K : L^2(N) \rightarrow H^{2t}(N)$ is compact and $B^{\text{adj}} : (H^{2t})^*(N) = H^{-2t}(N) \rightarrow (L^2)^*(N) = L^2(N)$ is the adjoint of the operator $B : L^2(N) \rightarrow H^{2t}(N)$ then

$$\text{index}(B + K) = \text{index}(B) = -\text{index}(B^{\text{adj}}).$$

Define $B_s : H^s(N) \rightarrow H^{s+2t}(N)$ as an extension of $B : C^\infty(N) \rightarrow C^\infty(N)$ and show that $\text{index}(B_s) = \text{index}(B_0)$ for all s . Define

$$P = (I - \Delta_g)^{s/2} B_s (I - \Delta_g)^{-s/2} : L^2(N) \rightarrow H^{2t}(N).$$

We can write $P = B_0 + K_1$ where $K_1 : L^2(N) \rightarrow H^{2t}(N)$ is compact. Now

$$\begin{aligned} \text{index}(B_0) &= \text{index}(P) \\ &= \text{index}(I - \Delta_g)^{s/2} + \text{index}(B_s) + \text{index}(I - \Delta_g)^{-s/2} \\ &= \text{index}(B_s). \end{aligned}$$

Because $B - B^{\text{adj}} : L^2(N) \rightarrow H^{2t}(N)$ is compact we can write

$$\begin{aligned} \text{index}(B : L^2(N) \rightarrow H^{2t}(N)) &= -\text{index}(B^{\text{adj}} : H^{-2t}(N) \rightarrow L^2(N)) \\ &= -\text{index}(B^{\text{adj}} : L^2(N) \rightarrow H^{2t}(N)) \\ &= -\text{index}(B : L^2(N) \rightarrow H^{2t}(N)) \end{aligned}$$

and hence we see that $\text{index}(B : L^2(N) \rightarrow H^{2t}(N)) = 0$. Using this, the knowledge that B is one-to-one and (3.2) we get

$$0 = \dim(\text{Ker } B) = \dim(\text{Coker } B)$$

which means that B is also onto. Thus we have shown that there exist $B^{-1} : H^{2t}(N) \rightarrow L^2(N)$.

Next we will examine Ψ DOs that depend on spectral variable $\lambda = (\alpha_0 \delta^k)^{-1}$. For the general theory see [42]. 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 for any compact set $K \subset \mathbb{R}^d$ there exist constants $C_{\alpha, \beta, K}$ 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 consider the pseudodifferential operators $A_\lambda : \mathcal{D}'(N) \rightarrow \mathcal{D}'(N)$ depending on the parameter λ . To define such operators, one considers local coordinates $Y : U \rightarrow \mathbb{R}^d$ of the manifold N , where we emphasize that the set $U \subset N$ does not need to be connected (see [42, section I.4.3]). A bounded linear operator $A_\lambda : \mathcal{D}'(N) \rightarrow \mathcal{D}'(N)$, depending on the parameter λ , is a pseudodifferential operator with spectral variable λ if for any local coordinates $Y : U \rightarrow \mathbb{R}^d$ of manifold N , $U \subset N$, there is a symbol $a \in S_p^m(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}_+)$ such that for $u \in C_0^\infty(U)$ we have

$$(A_\lambda u)(Y^{-1}(x)) = \int_{V \times \mathbb{R}^d} e^{i(x-y) \cdot \xi} a(x, \xi, \lambda) u(Y^{-1}(y)) dy d\xi, \quad x \in V,$$

where $V = Y(U) \subset \mathbb{R}^d$. In this case we will write

$$A_\lambda \in \Psi_p^m(N, \mathbb{R}_+),$$

and say that in local coordinates $Y : U \rightarrow \mathbb{R}^d$ the operator A has the symbol $a(x, \xi, \lambda) \in S_p^m(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}_+)$. If for all compact sets $K \subset \mathbb{R}^d$ 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

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

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

We want to prove that

$$F_\lambda = (A^*A)^{-1}(I - \Delta)^r + \lambda I$$

is invertible. Operator $F_\lambda \in \Psi^{2(t+r)}(N)$ is elliptic since $(A^*A)^{-1}(I - \Delta)^r \in \Psi^{2(t+r)}(N)$ is elliptic and $\lambda I \in \Psi^0(N)$. Denote $Q = (A^*A)^{-1}(I - \Delta)^r$ and its symbol $q(x, \xi) \in S^{2(t+r)}(N)$. Then for the symbol $\sigma(F_\lambda)(x, \xi) = q(x, \xi) + \lambda$ of the operator F_λ we have in compact subsets K of any local coordinates

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

By ([42] theorem 9.2.) there exist $R > 0$ such that for $|\lambda| \geq R$ the operator $F_\lambda \in H\Psi_{2(t+r)}^{2(t+r)}(N, \mathbb{R}_+)$ is invertible with

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

Now we have shown that the operator T^{-1} can be rewritten

$$T^{-1} = \lambda((A^*A)^{-1}(I - \Delta)^r + \lambda)^{-1}(A^*A)^{-1} \quad (3.3)$$

where $\lambda = (\alpha_0 \delta^\kappa)^{-1}$.

We denote by $\|F_\lambda\|_{s, s-\ell}$ the norm of $F_\lambda : H^s(N) \rightarrow H^{s-\ell}(N)$ where $s, \ell \in \mathbb{R}$. We have the following norm estimates for $F_\lambda \in \Psi_p^m(N, \mathbb{R}_+)$ when $\ell \geq m$ and λ large enough

$$\|F_\lambda\|_{s, s-\ell} \leq C_{s, l} (1 + |\lambda|^{1/p})^m, \quad \text{if } \ell \geq 0 \quad (3.4)$$

$$\|F_\lambda\|_{s, s-\ell} \leq C_{s, l} (1 + |\lambda|^{1/p})^{-(\ell-m)}, \quad \text{if } \ell \leq 0. \quad (3.5)$$

We can rewrite (3.1)

$$T_{\alpha(\delta)}(m_\delta) = u - \alpha T^{-1}(I - \Delta)^r u + \alpha_0^{-1} \delta^{1-\kappa} F_\lambda^{-1} (A^*A)^{-1} A^* \varepsilon. \quad (3.6)$$

For the third term on the right-hand side of (3.6) we have $(A^*A)^{-1} A^* : H^s(N) \rightarrow H^{\tilde{s}}(N)$, $\tilde{s} = s - t < -d/2$ and hence we get

$$\begin{aligned} \|F_\lambda^{-1} (A^*A)^{-1} A^* \varepsilon\|_{H^{s_1}} &= \|(I - \Delta_g)^{\frac{1}{2}(s_1 - \tilde{s})} F_\lambda^{-1} (A^*A)^{-1} A^* \varepsilon\|_{H^{\tilde{s}}} \\ &\leq \|(I - \Delta_g)^{\frac{1}{2}(s_1 - \tilde{s})} F_\lambda^{-1}\|_{\tilde{s}, s_1} \|(A^*A)^{-1} A^* \varepsilon\|_{H^{\tilde{s}}} \\ &\leq C \|(I - \Delta_g)^{\frac{1}{2}(s_1 - \tilde{s})} F_\lambda^{-1}\|_{\tilde{s}, s_1}. \end{aligned}$$

Define $E_\lambda = (I - \Delta_g)^{\frac{1}{2}(s_1 - \tilde{s})} F_\lambda^{-1} \in \Psi_p^m(N, \mathbb{R}_+)$, where $m = -2(t+r) + s_1 - \tilde{s}$ and $p = 2(t+r)$.

First we study the case when $s_1 \leq \tilde{s} = s - t < -d/2$. Inequality (3.4) gives us the norm estimate

$$\|E_\lambda\|_{\tilde{s}, \tilde{s}-\ell} \leq C(1 + |\lambda|^{1/p})^m = C(1 + \delta^{-\kappa/p})^m,$$

where $\ell = \tilde{s} - s_1 \geq 0$. Now clearly $\ell \geq m = -2(t+r) - \ell$. Because we want $\delta^{1-\kappa} \|F_\lambda^{-1} (A^*A)^{-1} A^* \varepsilon\|_{H^{s_1}} \rightarrow 0$ when $\delta \rightarrow 0$ we have to require that

$$-\frac{\kappa m}{p} = \kappa \left(1 + \frac{\ell}{2(t+r)} \right) > \kappa - 1.$$

Above is equivalent to

$$\frac{\ell}{2(t+r)} > -\frac{1}{\kappa}$$

which is true for all $\kappa > 1$, $t > 0$ and $r, \ell \geq 0$.

When $\tilde{s} \leq s_1$ we can use (3.5)

$$\|E_\lambda\|_{\tilde{s}, \tilde{s}-\ell} \leq C(1 + |\lambda|^{1/p})^{-(\ell-m)} = C(1 + \delta^{-\kappa/p})^{-(\ell-m)},$$

where $0 \geq \tilde{s} - s_1 = \ell \geq m = -2(t+r) - \ell$ if $-(t+r) \leq \ell \leq 0$. For convergence we need

$$\frac{\kappa(\ell - m)}{p} = \kappa \left(1 + \frac{\ell}{t+r} \right) > \kappa - 1$$

that is $0 \geq \ell > -(t+r)/\kappa$. For s_1 we get $s - t \leq s_1 \leq s - t - (t+r)/\kappa$.

When $r > 0$ the convergence of $\alpha T^{-1}(I - \Delta)^r u$ could be shown the same way as in the previous example. Next we will show the convergence also in the case $r = 0$ and improve the convergence rate by proving the convergence in H^ζ instead of L^2 .

Assume that $r \geq 0$ and denote $\zeta = -r - \theta \geq s_1$. We need to find such $\eta \geq 0$ and $\gamma \geq 0$ that $\gamma + \eta = 1$ and $t\gamma - r\eta - \theta/2 = 0$. Define $\eta = (2t - \theta)/2(t+r)$ and $\gamma = (2r + \theta)/2(t+r)$, where $\theta \leq 2t$. Using the inequalities $T = A^*A + \alpha(I - \Delta)^r \geq \alpha(I - \Delta)^r$ and $T \geq A^*A \geq c_1(I - \Delta)^{-t}$ we get

$$\begin{aligned} \|\alpha T^{-1}(I - \Delta)^r u\|_{H^\zeta} &\leq \alpha \|(c_1(I - \Delta)^{-t})^{-\gamma} (\alpha(I - \Delta)^r)^{-\eta} (I - \Delta)^{r - \frac{t}{2} - \frac{\theta}{2}} u\|_{L^2} \\ &\leq c_1^{-\gamma} \alpha^{1-\eta} \|(I - \Delta)^{t\gamma - r\eta - \frac{\theta}{2}} (I - \Delta)^{\frac{t}{2}} u\|_{L^2} \\ &= c_1^{-\gamma} \delta^{\frac{\kappa(r-\zeta)}{2(t+r)}} \|u\|_{H^r} \end{aligned}$$

where $\zeta = \max\{s_1, -r - 2t\}$.

Adding the above results together we can prove theorem 1.

Proof of theorem 1. The convergence is immediate consequence of the above results. Now when $s_1 \leq \tilde{s}$ we have

$$\begin{aligned} \|T_{\alpha(\delta)}(m_\delta) - u\|_{H^{s_1}} &\leq \alpha \|T^{-1}(I - \Delta)u\|_{H^\zeta} + \alpha_0^{-1} \delta^{1-\kappa} \|F_\lambda^{-1}(A^*A)^{-1} A^* \varepsilon\|_{H^{s_1}} \\ &\leq C_1 \delta^{\frac{\kappa(r-\zeta)}{2(t+r)}} + C_2 \delta^{1-\kappa + \kappa(1 + \frac{\ell}{2(t+r)})} \\ &\leq C_3 \max \left\{ \delta^{\frac{\kappa(r-\zeta)}{2(t+r)}}, \delta^{1 + \frac{\kappa(s-t-s_1)}{2(t+r)}} \right\}. \end{aligned}$$

If $\tilde{s} \leq s_1 \leq s - t + (t+r)/\kappa$ we get

$$\begin{aligned} \|T_{\alpha(\delta)}(m_\delta) - u\|_{H^{s_1}} &\leq \alpha \|T^{-1}(I - \Delta)u\|_{H^\zeta} + \alpha_0^{-1} \delta^{1-\kappa} \|F_\lambda^{-1}(A^*A)^{-1} A^* \varepsilon\|_{H^{s_1}} \\ &\leq C_1 \delta^{\frac{\kappa(r-\zeta)}{2(t+r)}} + C_2 \delta^{1-\kappa + \kappa(1 + \frac{\ell}{t+r})} \\ &\leq C_3 \max \left\{ \delta^{\frac{\kappa(r-\zeta)}{2(t+r)}}, \delta^{1 + \frac{\kappa(s-t-s_1)}{t+r}} \right\}. \end{aligned}$$

Above $\zeta = \max\{s_1, -r - 2t\}$. This means that the noise term is dominating when $s_1 \geq s - r - 2t$. This proves theorem 1. \square

4. A model problem: one-dimensional deblurring

We consider a simple inverse problem to give flavour of results for the reader. Let \mathbb{T}^2 be the two-dimensional torus constructed by identifying parallel sides of the square $D = (0, 1)^2 \subset \mathbb{R}^2$; we model periodic images as elements of function spaces over \mathbb{T}^2 . The continuum model is $m = Au + \varepsilon$ with convolution operator A defined by

$$Au(x) = \int_{\mathbb{T}^2} \Phi(x - y)u(y) dy, \tag{4.1}$$

where $\Phi \in C(\mathbb{T}^2)$ is a point spread function that is given by the Schwartz kernel of an elliptic pseudodifferential operator of the order $-t < -2$.

4.1. Divergence in H^1 -norm

Let us return to the translation-invariant case where we wrote the regularized solution in the form

$$T_{\alpha(\delta)}(m_\delta) = v_\delta + w_\delta$$

where

$$v_\delta = (A^*A + \alpha_0\delta^2(I - \Delta))^{-1}A^*Au$$

and

$$w_\delta = (A^*A + \alpha_0\delta^2(I - \Delta))^{-1}A^*(\delta\varepsilon).$$

Now the Fourier transform of w_δ is

$$\widehat{w}_\delta(n) = \frac{1}{z_\delta(n)} \overline{a(n)} \widehat{\varepsilon}(n) \delta.$$

Denote $I(\delta) = \{n \mid c_0\delta^2(1+n^2) \leq |a(n)|^2 \leq c_1\delta^2(1+n^2)\}$. We get

$$\begin{aligned} \|w_\delta\|_{H^1}^2 &\geq \sum_{I(\delta)} (1+n^2) \left| \frac{\overline{a(n)}\delta}{|a(n)|^2 + \alpha_0\delta^2(1+n^2)} \widehat{\varepsilon}(n) \right|^2 \\ &\geq \sum_{I(\delta)} \frac{|a(n)|^2}{|a(n)|^2 + \alpha_0\delta^2(1+n^2)} \cdot \frac{(1+n^2)\delta^2}{|a(n)|^2 + \alpha_0\delta^2(1+n^2)} |\widehat{\varepsilon}(n)|^2 \\ &\geq \sum_{I(\delta)} \frac{1}{1 + \frac{\alpha_0}{c_0}} \cdot \frac{1}{c_1 + \alpha_0} |\widehat{\varepsilon}(n)|^2. \end{aligned}$$

We can choose c_0, c_1 so that $I(\delta) \neq \emptyset$ for all $0 < \delta < \delta_0$. Now there exist $n(\delta) \in I(\delta)$ and $n(\delta)$ goes through all $\{n \in \mathbb{N} \mid n \geq n_0\}$ when $\delta \rightarrow 0$. We see that

$$\limsup_{\delta \rightarrow 0} \|w_\delta\|_{H^1}^2 \geq \limsup_{\delta \rightarrow 0} c_2 |\widehat{\varepsilon}(n(\delta))|^2 \geq c_2$$

almost surely since ε is white noise. Thus the solution $T_{\alpha(\delta)}(m_\delta)$ does not converge in H^1 .

4.2. Computational results

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 discretization is refined, i.e., when $n \rightarrow \infty$ or $k \rightarrow \infty$. This is the very reason why regularization is needed in the (numerical) solutions of the inverse problems.

Next we demonstrate the above results numerically and consider one-dimensional deblurring problem on the torus $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$,

$$m = Au + \delta\varepsilon,$$

where $u \in H^1(\mathbb{T}^1)$ is the following piecewise linear function (see figure 1):

$$u = \begin{cases} 0 & \text{when } 0 < x < 0.3 \text{ or } 0.7 < x < 1 \\ 10x - 3 & \text{when } 0.3 < x < 0.4 \\ 1 & \text{when } 0.4 \leq x \leq 0.6 \\ -10x + 7 & \text{when } 0.6 < x < 0.7, \end{cases}$$

$\varepsilon \in H^s$, $s < -1/2$ is white noise and A is a two times smoothing operator

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

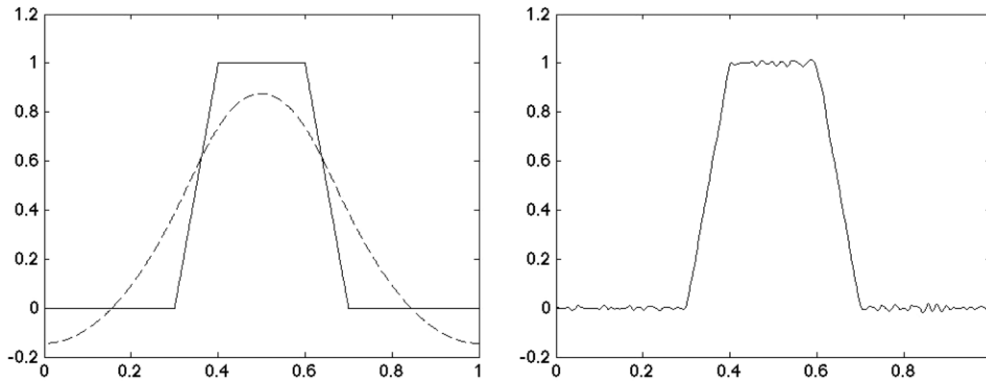


Figure 1. On the left the original piecewise linear function u (solid line) and the noiseless data $m = Au$ (dashed line). On the right regularized solution u_δ when $\delta = 3, 5 \times 10^{-5}$.

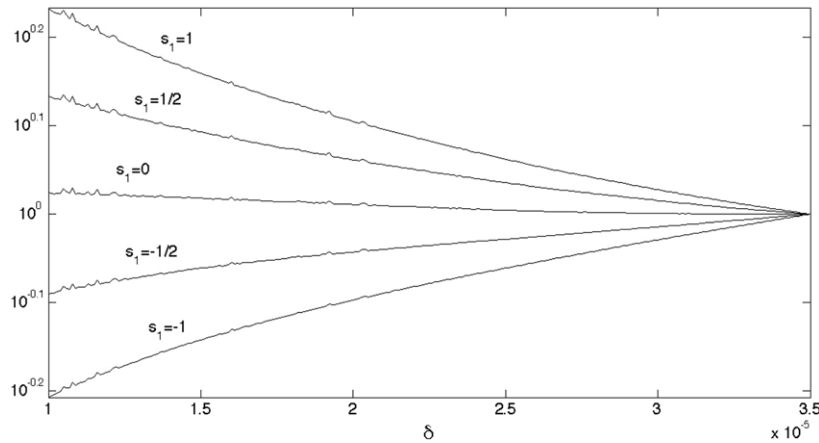


Figure 2. Normalized errors $c(s_1)\|u - u_\delta\|_{H^{s_1}(\mathbb{T}^1)}$ in logarithmic scale with different values of s_1 . We observe that u_δ converge to u at least when $s_1 \leq -1/2$.

Now 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.

We assume $\alpha = \delta^{5/2}$ and thus the regularized solution is

$$u_\delta = T_{\alpha(\delta)}(m_\delta) = (A^*A + \delta^{5/2}(I - \Delta)^r)^{-1}A^*m.$$

Now theorem 1 gives us

$$\lim_{\delta \rightarrow 0} \|u - u_\delta\|_{H^{s_1}} = 0$$

when $s_1 < s - t + (t + r)/\kappa < -13/10$.

We know that $u, u_\delta \in H^1$ for all $\delta > 0$ and are interested to know what happens in H^1 when $\delta \rightarrow 0$. From figure 2 we can see that u_δ converges to u in H^{s_1} at least when $s_1 < -1/2$. On the other hand even though both functions belong to space H^1 we do not have convergence there that is $u_\delta \not\rightarrow u$ in H^1 .

5. Convergence of the finite-dimensional minimization problems

In this section we consider the convergence of the finite-dimensional minimization problems (1.4) to the ideal, infinite-dimensional model (1.12).

Below, we consider the case when P_k is given by formula (1.8) where ϕ_j are the eigenfunctions of the Laplace operator of N such that $(\phi_j)_{j=1}^\infty$ is an orthogonal basis in $L^2(N)$. Then $P_k : H^s(N) \rightarrow \mathbb{R}^k$ is a bounded linear map for all $s \in \mathbb{R}$.

We will first consider the relation of the noise models (1.6) and (1.7). Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. Assume that $\mathcal{E} : \Omega \rightarrow \mathcal{D}'(Y)$ is a Gaussian random generalized function (see [4]) with a covariance operator $C_{\mathcal{E}}$ that can be extended to a bounded map $C_{\mathcal{E}} : H^{\tilde{s}}(N) \rightarrow H^{-\tilde{s}}(N)$ with some $\tilde{s} \in \mathbb{R}$. Let us define $\tilde{\mathcal{E}} = (I - \Delta)^{p/2 + \tilde{s}/2} \mathcal{E}$. Then the covariance operator of $\tilde{\mathcal{E}}$ is $C_{\tilde{\mathcal{E}}} = (I - \Delta)^{-p/2 - \tilde{s}/2} C_{\mathcal{E}} (I - \Delta)^{-p/2 - \tilde{s}/2}$. Weyl's theorem implies that the eigenvalues of $-\Delta$ have the asymptotics $\lambda_j(-\Delta) = c j^{2/d} (1 + O(j^{-1}))$ and thus $(I - \Delta)^{-p/2}$ is a Hilbert–Schmidt operator, that is $(I - \Delta)^{-p/2} \in \mathfrak{S}_2$, when $p > d/2$. On the other hand $(I - \Delta)^{-\tilde{s}/2} C_{\mathcal{E}} (I - \Delta)^{-\tilde{s}/2} : L^2 \rightarrow L^2$ since $C_{\mathcal{E}} : H^{\tilde{s}} \rightarrow H^{-\tilde{s}}$ is bounded and hence we can conclude that $C_{\tilde{\mathcal{E}}}$ is a trace class operator which implies that \mathcal{E} is almost surely a $H^{-p-\tilde{s}}(N)$ -valued random function.

Under the above assumptions, consider the \mathbb{R}^k -valued random variable $\tilde{\mathbf{E}}^{(k)} = P_k \mathcal{E}$. It has the covariance operator $P_k C_{\mathcal{E}} P_k^*$. Thus the random variable $\tilde{\mathbf{E}}^{(k)}$ has the distribution $N(0, \delta^2 I)$ for all $k \in \mathbb{Z}_+$ if and only if $C_W = \delta^2 I$, or equivalently, \mathcal{E} has the same distribution as $W\delta$, where $W : \Omega \rightarrow \mathcal{D}'(Y)$ is the normalized Gaussian white noise given by (1.7). This is the reason behind our assumption that in the ideal, infinite-dimensional model we have noise = $\varepsilon(x)\delta$ where ε is a realization of the normalized Gaussian white noise W .

To study the convergence of the finite-dimensional minimization problems (1.4) to the infinite-dimensional problem (1.12) we use the Γ -convergence, see [1, 2, 33]. Let (Y, d_Y) be metric space and τ be the topology of Y induced by the metric d_Y . Below we will in particular consider the case when $Y \subset X$ is a closed bounded subset of Banach space X for which the dual space X' is separable. Then the weak topology of X induces a topology τ for the subset Y that is metrizable, that is, induced by some metric $d_1 : Y \times Y \rightarrow \mathbb{R}$, see e.g. [33, proposition 8.7]. Note that the metric d_1 is not necessarily the metric induced by the norm of X . On following definition, see e.g. [2, definition 2.1.7] or [33, proposition 8.1].

Definition 3. Let $Y \subset X$ be a closed bounded subset of Banach space X for which the dual space X' is separable. We say that $F_j : Y \rightarrow \mathbb{R} \cup \{\infty\}$ Γ -converges to $F : Y \rightarrow \mathbb{R} \cup \{\infty\}$ with the topology τ and denote $F = \Gamma\text{-}\lim_{j \rightarrow \infty} F_j$ if

- (i) for every $u \in Y$ and for every sequence u_j τ -converging to u in Y we have $F(u) \leq \liminf_{j \rightarrow \infty} F_j(u_j)$,
- (ii) for every $u \in Y$ there exists a sequence u_j τ -converging to u in Y such that $F(u) \geq \limsup_{j \rightarrow \infty} F_j(u_j)$.

We need also the concept of equicoercivity, see [2, definition 2.1.8].

Definition 4. Let $Y \subset X$ be a closed bounded subset of Banach space X for which the dual space X' is separable. We call a sequence of functionals $F_j : Y \rightarrow \mathbb{R} \cup \{\infty\}$, $j \in \mathbb{Z}_+$, equicoercive in topology τ if for every $t \geq 0$ there exists a compact set $K_t \subset Y$ such that $\{u \in Y \mid F_j(u) \leq t\} \subset K_t$ for all $j \in \mathbb{Z}_+$.

Using these definitions, we return to the setting of the problem given in section 1 where N is a d -dimensional compact closed manifold, $A \in \Psi^{-t}$ is a pseudodifferential operator, and $m = m_\delta$ is the measurement (1.15).

Let us consider the finite-dimensional minimization problems analogous to (1.4), that are given by

$$T_{\alpha;n,k}(m) := \arg \min_{u \in X_n} F_{n,k}(u), \quad (5.1)$$

where $n, k \in \mathbb{Z}_+$ and $F_{n,k} : H^r(N) \rightarrow \mathbb{R} \cup \{\infty\}$,

$$F_{n,k}(u) = \|P_k A u - P_k m\|_2^2 + \alpha \|u\|_{H^r(N)}^2, \quad \text{for } u \in X_n, \quad (5.2)$$

and $F_{n,k}(u) = \infty$ for $u \notin X_n$, where $X_n \subset H^r(N)$ is a n -dimensional subspace. Also, let $G : H^r(N) \rightarrow \mathbb{R}$ be

$$G(u) = \|A u\|_{L^2(N)}^2 - 2\langle m, A u \rangle + \alpha \|u\|_{H^r(N)}^2, \quad \text{for } u \in H^r(N).$$

Let $Y = \{u \in H^r(N); \|u\|_{H^r(N)} \leq C_0\}$, where

$$C_0 > 2\alpha^{-1} \max(\|A^* m\|_{H^{-r}(N)}, \sup_{k \in \mathbb{Z}_+} \|(P_k A)^* m\|_{H^{-r}(N)})$$

so that $G(u) > G(0) = 0$ and $F_{n,k}(u) > F_{n,k}(0) = \|P_k \varepsilon\|_2^2$ for all $\|u\|_{H^r(N)} > C_0$. Thus functions $F_{n,k} : H^r(N) \rightarrow \mathbb{R} \cup \{\infty\}$ and $G : H^r(N) \rightarrow \mathbb{R} \cup \{\infty\}$ obtain their minimal values in Y . We endow Y with the relative topology determined by the weak topology of $H^r(N)$.

Proposition 5. *Let the assumptions of theorem 1 hold, in particular, let $\varepsilon \in H^s(N)$ with $s < -d/2$ and $m = m_\delta$ be the measurement given by (1.15). Moreover, assume that $X_n \subset X_{n+1}$ and $\cup_{n=1}^\infty X_n$ is a dense subset of $H^r(N)$ and let $c_k = \|P_k m\|_2^2$. Then the functions $G_{n,k} : Y \rightarrow \mathbb{R} \cup \{\infty\}$,*

$$G_{n,k}(u) = F_{n,k}(u) - c_k,$$

converge to $G : Y \rightarrow \mathbb{R}$ as $n, k \rightarrow \infty$ in sense of the Γ -convergence with respect to the topology of Y . Moreover, the minimizers $T_{\alpha;n,k}(m)$ of $F_{n,k}$ converge to the unique minimizer $T_\alpha(m)$ of $G : H^r(N) \rightarrow \mathbb{R}$ in the weak topology of $H^r(N)$ as $n, k \rightarrow \infty$.

Proof. Let $u \in Y$ and let $u_{n,k} \in Y$ be a sequence that converge to u weakly in $H^r(N)$ as $n, k \rightarrow \infty$. As the linear operator $A : H^r(N) \rightarrow L^2(N)$ is a compact operator, $P_k A u_{n,k}$ converge to $A u$ in the strong topology of $L^2(N)$ as $n, k \rightarrow \infty$. Moreover, the map $u \mapsto \|u\|_{H^r(N)}$ is lower a semi-continuous function in Y . These facts imply that the property (i) in definition 3 holds.

Let Q_n be orthogonal projectors in $H^r(N)$ onto the subspace X_n . Let $u \in Y$, and define for $n, k \in \mathbb{Z}_+$ $u_{n,k} = Q_n u$. Then

$$G_{n,k}(u_{n,k}) = \|P_k A Q_n u - P_k m\|_2^2 + \alpha \|Q_n u\|_{H^r(N)}^2 - c_k$$

converge to $G(u)$ as $n, k \rightarrow \infty$, and we see that the property (ii) in definition 3 is valid. Thus $G_{n,k}$ Γ -converge to G as $n, k \rightarrow \infty$.

Since all closed subsets of Y are compact, we see that $\{G_{n,k} : Y \rightarrow \mathbb{R} \cup \{\infty\}; n, k \in \mathbb{Z}_+\}$ is an equicoercive family of functions. Moreover, the functions $G_{n,k} : Y \rightarrow \mathbb{R} \cup \{\infty\}$ and $F : Y \rightarrow \mathbb{R}$ have unique minimizers and the minimizer of $F_{n,k} : Y \rightarrow \mathbb{R} \cup \{\infty\}$ is equal to the minimizers $T_{\alpha;n,k}(m)$ and finally, the minimizer of $G : Y \rightarrow \mathbb{R}$ is equal to $T_\alpha(m)$. Thus by [33, corollary 7.24], the minimizers $T_{\alpha;n,k}(m)$ of the functions $G_{n,k}$ converge weakly in $H^r(N)$ to the minimizer of G as $n, k \rightarrow \infty$. \square

6. Conclusion

We discuss above finite-dimensional linear models of indirect measurement corrupted by white Gaussian noise. Such models are used in countless practical inverse problems. It is desirable to connect these discrete models to an infinite-dimensional limit model. Such a connection can provide, for instance, error analysis for numerical inversion and computational speed-ups based on robust switching between different discretizations related to multigrid methods.

The focus of our analysis is the apparent paradox arising from the (almost surely) infinite L^2 -norm of the natural limit of white Gaussian noise in \mathbb{R}^n as $n \rightarrow \infty$. We show how to build a rigorous theory removing this paradox, and we explain how to take this into account in discrete inverse problems using appropriate Sobolev space norms.

Proposition 5 shows that the infinite-dimensional minimization problem (1.14) is the natural limit of the finite-dimensional minimization problems (5.1). Therefore, when the measured data is corrupted by white Gaussian noise, despite the fact that the realizations of the white noise are almost surely not L^2 -functions, the inner product associated to the L^2 -norm is appropriate for data fidelity terms when the inverse problems are solved using Tikhonov regularization. Moreover, our results show how the regularization parameters can be chosen to obtain converging results when the noise amplitude goes to zero.

Our results pave the way to numerical analysis of Tikhonov regularization based on fruitful interplay between discrete and continuous models.

Acknowledgments

This work was supported by the Finnish Centre of Excellence in Inverse Problems Research 2012-2017 (Academy of Finland CoE-project 250215). In addition, HK and ML were supported by Academy of Finland, project 141104, and SS was supported by Academy of Finland, project 141094.

References

- [1] Attouch H 1984 *Variational Convergence for Functions and Operators (Applicable Mathematics Series vol 14)* (Boston, MA: Pitman) p 423
- [2] Aubert G and Kornprobst P 2006 *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations vol 147* (Berlin: Springer)
- [3] Bissantz N, Hohage T and Munk A 2004 Consistency and rates of convergence of nonlinear Tikhonov regularisation with random noise *Inverse Problems* **20** 1773–89
- [4] Bogachev V 1998 *Gaussian Measures (Mathematical Surveys and Monographs vol 62)* (Providence, RI: American Mathematical Society) p 433
- [5] Burger M and Neubauer A 2003 Analysis of Tikhonov regularization for function approximation by neural networks *Neural Netw.* **16** 79–90
- [6] Calvetti D and Somersalo E 2007 A Gaussian hypermodel to recover blocky objects *Inverse Problems* **23** 733–54
- [7] Egger H 2008 Regularization of inverse problems with large noise *J. Phys.: Conf. Ser.* **124** 012022
- [8] Eggermont P P B, LaRiccia V N and Nashed M Z 2009 On weakly bounded noise in ill-posed problems *Inverse Problems* **25** 115018–32
- [9] Eggermont P P B, LaRiccia V N and Nashed M Z 2012 Moment discretization for ill-posed problems with discrete weakly bounded noise *Int. J. Geomathematics* **3** 155–78
- [10] Eggermont P P B, LaRiccia V N and Nashed M Z 2010 Noise models for ill-posed problems *Handbook of Geomathematics* (Berlin: Springer) pp 739–62
- [11] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems vol 375* (Berlin: Springer)

- [12] Franklin J N 1970 Well-posed stochastic extensions of ill-posed linear problems *J. Math. Anal. Appl.* **31** 682–716
- [13] Fitzpatrick B 1991 Bayesian analysis in inverse problems *Inverse Problems* **7** 675–702
- [14] Groetsch C W 1984 *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind* vol 105 (Boston, MA: Pitman)
- [15] Hanson K 1987 Bayesian and related methods in image reconstruction from incomplete data *Image Recovery: Theory and Applications* (Orlando FL: Academic) 79–125
- [16] Hohage T and Pricop M 2008 Nonlinear Tikhonov regularization in Hilbert scales for inverse boundary value problems with random noise *Inverse Problems Imaging* **2** 271–90
- [17] Hohage T and Werner F 2013 Iteratively regularized Newton-type methods for general data fidelity functionals and applications to Poisson data *Numer. Math.* **123** 745–9
- [18] Hörmander L 2007 *The Analysis of Linear Partial Differential Operators III: Pseudo-Differential Operators* vol 274 (Berlin: Springer)
- [19] Kaipio J and Somersalo E 2005 *Statistical and Computational Inverse Problems* vol 160 (New York: Springer)
- [20] Kallenberg O 2002 *Foundations of Modern Probability Theory* (Berlin: Springer)
- [21] Kirsch A 2011 *An Introduction to the Mathematical Theory of Inverse Problems* vol 120 (Berlin: Springer)
- [22] Klann E, Maass P and Ramlau R 2006 Two-step regularization methods for linear inverse problems *J. Inverse Ill-posed Problems* **14** 583–607
- [23] Klann E and Ramlau R 2008 Regularization by fractional filter methods and data smoothing *Inverse Problems* **24** 025018
- [24] Kolehmainen V, Siltanen S, Järvenpää S, Kaipio J P, Koistinen P, Lassas M, Pirttilä J and Somersalo E 2003 Statistical inversion for X-ray tomography with few radiographs II: Application to dental radiology *Phys. Med. Biol.* **48** 1465–90
- [25] Lasanen S 2002 Discretizations of generalized random variables with applications to inverse problems *Ann. Acad. Sci. Fenn. Math. Diss.* **130** 64
- [26] Lasanen S 2012 Non-Gaussian statistical inverse problems. Part I: posterior distributions *Inverse Problems Imaging* **6** 215–66
- [27] Lasanen S 2012 Non-Gaussian statistical inverse problems. Part II: posterior convergence for approximated unknowns *Inverse Problems Imaging* **6** 267–87
- [28] Lassas M, Saksman S and Siltanen S 2009 Discretization invariant Bayesian inversion and Besov space priors *Inverse Problems Imaging* **3** 87–122
- [29] Lassas M and Siltanen S 2004 Can one use total variation prior for edge preserving Bayesian inversion? *Inverse Problems* **20** 1537–64
- [30] Lehtinen M, Päivärinta L and Somersalo E 1989 Linear inverse problems for generalised random variables *Inverse Problems* **5** 599–612
- [31] Lunttila T, Nenonen J and Somersalo E 2003 Regularisation in cardiac source imaging *Functional Imaging and Modeling of the Heart* (Berlin: Springer) pp 101–10
- [32] Mair B A and Ruymgaart F H 1996 Statistical inverse estimation in Hilbert scales *SIAM J. Appl. Math.* **56** 1424–44
- [33] Maso G Dal 1993 *An Introduction to Γ -Convergence* vol 8 (Berlin: Springer)
- [34] Mathé P and Tautenhahn U 2011 Regularization under general noise assumptions *Inverse Problems* **27** 035016
- [35] Mathé P and Tautenhahn U 2011 Enhancing linear regularization to treat large noise *J. Inverse Ill-posed Problems* **19** 859–79
- [36] Mathé P and Pereverzev S V 2001 Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods *SIAM J. Numer. Anal.* **38** 1999–2021
- [37] Morozov V A and Nashed Z 1984 *Methods for Solving Incorrectly Posed Problems* (New York: Springer)
- [38] Phillips B L 1962 A technique for the numerical solution of certain integral equations of the first kind *J. Assoc. Comput. Mach.* **9** 84–97
- [39] Neubauer A and Pikkarainen H K 2008 Convergence results for the Bayesian inversion theory *J. Inverse Ill-posed Problems* **16** 601–13
- [40] Pikkarainen H 2006 State estimation approach to nonstationary inverse problems: discretisation error and filtering problem *Inverse Problems* **22** 365–79
- [41] Piiroinen P 2005 Statistical measurements, experiments and applications *Ann. Acad. Sci. Fenn. Math. Diss.* **143** 89

- [42] Shubin M A 2001 *Pseudodifferential Operators and Spectral Theory* (Berlin: Springer)
- [43] Siltanen S, Kolehmainen V, Järvenpää S, Kaipio J P, Koistinen P, Lassas M, Pirttilä J and Somersalo E 2003 Statistical inversion for x-ray tomography with few radiographs I: General theory *Phys. Med. Biol.* **48** 1437–63
- [44] Sudakov V N and Khalfin L A 1964 A statistical approach to the correctness of the problems of mathematical physics *Dokl. Akad. Nauk SSSR* **157** 1058–60 (in Russian)
- [45] Tarantola A 2005 *Inverse Problem Theory and Methods for Model Parameter Estimation* (Philadelphia: SIAM)
- [46] Tikhonov A N 1943 On the stability of inverse problems *Dokl. Akad. Nauk SSSR* **39** 195–8 (in Russian)
- [47] Tikhonov A N 1963 On the solution of incorrectly formulated problems and the regularization method *Sov. Math. Dokl.* **4** 1035–8 (in Russian)
- [48] Tikhonov A N 1995 *Numerical Methods for the Solution of Ill-Posed Problems* vol 328 (Berlin: Springer)