Posterior consistency and convergence rates for Bayesian inversion

Hanne Kekkonen

joint work with

Matti Lassas and Samuli Siltanen

Department of Mathematics and Statistics
University of Helsinki, Finland

June 2, 2015

The indirect measurement problem

Our starting point is the continuous linear measurement model

$$M = AU + \mathcal{E}\delta, \qquad \delta > 0 \tag{1}$$

- where M, U and $\mathcal{E}$ are treated as random variables.
- The unknown U takes values in $H^{-\tau}(N)$ with some $\tau \in \mathbb{R}$.
- We assume $\mathcal{E}$ to be Gaussian white noise taking values in $H^{-s}(N)$, $s > d/2$.

The unknown is treated as a random variable since we have only incomplete data of U.

Bayes formula combines data and a priori information

The inverse problem is to find an estimate for U if we are given a realisation m of the measurement M.

Bayes' formula for discrete problem

Bayes' formula gives us the posterior distribution $\pi(u_u \mid m_k)$:

$$\begin{aligned}\pi(u_n \mid m_k) &= C\pi_{pr}(u_n)\pi_{\varepsilon}(m_k \mid u_n)\\ &= C\exp\Big(-\frac{1}{2\delta^2}\|m_k - \mathbf{A}u_n\|_{\ell^2}^2 - \frac{1}{2}\|\mathbf{C}_U^{-1/2}u_n\|_{\ell^2}^2\Big).\end{aligned} \tag{2}$$

The result of Bayesian inversion is the posterior distribution, but typically one looks at MAP or CM estimate.

We don’t have Bayes’ formula for continuous problem

If we assume that that the noise takes values in L^2 the MAP-estimate of (2) Γ-converges to the following infinite-dimensional minimisation problem:

$$\underset{u \in H^r}{\operatorname{argmin}} \Big\{ \frac{1}{2\delta^2} \|m - Au\|_{L^2}^2 + \frac{1}{2} \|C_U^{-1/2} u\|_{L^2}^2 \Big\}. \tag{3}$$

Now if we think that the above is a MAP estimate of a Bayesian problem we have to assume that U has formally the following distribution

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

Above we assume that C_U is a $2r$ times smoothing covariance operator.

Does white noise belong to L^2?

Formally

$$\varepsilon = \sum_{j=-\infty}^{\infty} \langle \varepsilon, \psi_j \rangle \psi_j$$

where ψ_j form an orthonormal basis for L^2. The Fourier coefficients of white noise satisfy $\langle \varepsilon, e_k \rangle \sim N(0,1)$, where $e_k(t) = e^{ikt}$. Hence

$$\|\varepsilon\|_2^2 = \sum_{k=-\infty}^{\infty} |\langle \varepsilon, e_k \rangle|^2 < \infty \qquad \text{with probability zero.}$$

For the white noise we have

i) $\varepsilon \in L^2$ with probability zero,

ii) $\varepsilon \in H^{-s}$, $s > d/2$, with probability one.

"The white noise paradox"

If we are working on a discrete world $\|\varepsilon_k\|_{\ell^2} < \infty$ with all $k \in \mathbb{R}$. Hence the minimisation problem

$$u_n^\delta = \underset{u}{\operatorname{argmin}} \left\{ \|\mathbf{A}u_n - m_k\|_{\ell^2}^2 + \alpha \|\mathbf{C}_U^{-1/2} u_n\|_{\ell^2}^2 \right\}$$

is well defined. However we know that

$$\lim_{k\to\infty} \|\varepsilon_k\|_{\ell^2} = \infty.$$

The goal is to build a rigorous theory removing the apparent paradox arising from the infinite L^2-norm of the natural limit of white Gaussian noise in $\mathbb{R}^k$ as $k \to \infty$.

We can define new MAP estimate by omitting the constant term $\|m\|_{L^2}^2$

When $m - Au \in L^2$ we can write

$$\|m - Au\|_{L^2}^2 = \|Au\|_{L^2}^2 - 2\langle m, Au\rangle_{L^2} + \|m\|_{L^2}^2.$$

Now omitting the constant term $\|m\|_{L^2}^2$ in (3) we get a new well defined minimisation problem

$$u_\delta = \arg\min_{u \in H^r} \left\{ \|Au\|_{L^2}^2 - 2\langle m, Au\rangle + \delta^2 \|C_U^{-1/2} u\|_{L^2}^2 \right\}.$$

The solution to the problem above is

$$u_\delta = \left(A^* A + \delta^2 C_U^{-1} \right)^{-1} (A^* m)$$

where A is a pseudodifferential operator.

Does omitting $\|m\|_{L^2}^2 = \infty$ cause any troubles?

Example

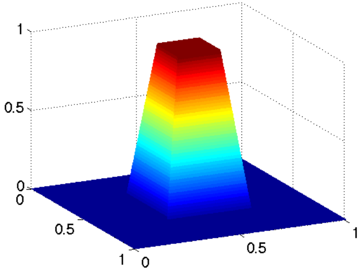
We consider the problem

$$m = Au + \varepsilon\delta = \int \Phi(\cdot - y)u(y)dy + \varepsilon\delta$$

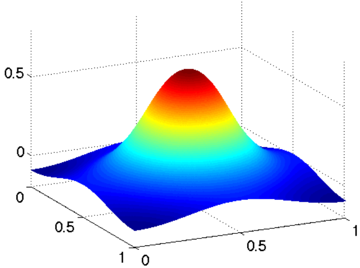
where $u \in H^1$ is a piecewise linear function, ε is white noise and

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

We have $u \in H^1$ and $u_\delta \in H^1$ for all $\delta > 0$ so does $u_\delta \to u$ in H^1 when $\delta \to 0$?



The unknown function u.



Noiseless data $m = Au$

Solution u_δ does not converge to u in H^1

We are interested in knowing what happens to the regularised solution u_δ in different Sobolev spaces when $\delta \to 0$.

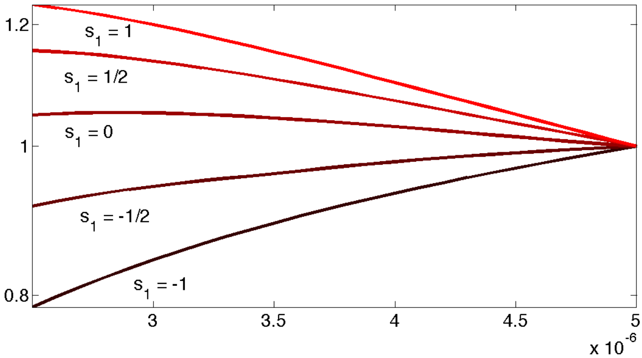


Figure: Normalised errors $c(s_1)\|u - u_\delta\|_{H^{s_1}(\mathbb{T}^1)}$ in logarithmic scale with different values of s_1. We observe that $\lim_{\delta \to 0} \|u - u_\delta\|_{H^1(\mathbb{T}^1)} \neq 0$.

Why are we interested in continuous white noise?

It is important to be able to connect discrete models to their infinite-dimensional limit models.

- In practice we do not solve the continuous problem but its discretisation.
- Discrete white noise is used in many practical inverse problems as a noise model.
- If the discrete model is an orthogonal projection of the continuous model to a finite dimensional subspace it guarantees that we can switch consistently between different discretisations which is important for e.g. multigrid methods.

Brief literature review

1989 Lehtinen, Päivärinta and Somersalo

The conditional distribution exists in spaces of generalised functions

2000 Ghosal, Ghosh, and Van Der Vaart

Posterior consistency and convergence rates of posterior distributions

2001 Shen and Wasserman

Posterior consistency and convergence rates of posterior distributions

2011 Knapik, Van Der Vaart and Van Zanten

Posterior contraction results with diagonalisable operators

2013 Agapiou, Larsson and Stuart

Posterior contraction results with Gaussian priors

2013 Dashti, Law, Stuart and Voss

Consistency of MAP estimators in Bayesian inverse problems

Crash course to generalised random variables

White noise $\mathcal{E}$ can be considered as a measurable map $\mathcal{E} : \Omega \to \mathcal{D}'(N)$ where Ω is a probability space. Then white noise $\mathcal{E}(y, \omega)$ is a random generalised function for which:

- pairings $\langle \mathcal{E}, \phi \rangle_{\mathcal{D}' \times \mathcal{D}}$ are Gaussian random variables for all test functions $\phi \in \mathcal{D} = C_0^\infty(N)$,
- we have $\mathbb{E}\mathcal{E} = 0$ and
- the covariance operator $C_\mathcal{E} = I$ where we define

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

Below we will write $\mathcal{E} \sim N(0, I)$ as shorthand.

Rigorous way of gaining conditional mean estimate

Assume that the unknown and the white noise are independent and Gaussian

$$U \sim N(0, C_U), \qquad \mathcal{E} \sim N(0, I).$$

Then the posterior distribution, that is the conditional distribution of $u|m$, is Gaussian and has the mean

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

This is equivalent to the MAP estimate defined above.

Note that in Gaussian case the MAP estimate coincide almost surely with the CM estimate.

Simple example in $\mathbb{T}^1$

Next we assume that $U, \mathcal{E} \sim N(0, I)$. We know that the realisations $u, \varepsilon \in H^{-s}$, $s > 1/2$ a.s. The unknown U has the formal distribution

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

Solving the CM/MAP estimate is linked to solving the minimisation problem

$$u_\delta = \underset{u \in L^2}{\operatorname{argmin}} \left\{ \|Au\|_{L^2}^2 - 2\langle Au, m \rangle + \delta^2 \|u\|_{L^2}^2 \right\}.$$

That is we are looking for approximation in L^2 even though the realisations of U are in L^2 with probability zero.

What can we say in a general case?

Let C_U be $2r$ times smoothing, self-adjoint, injective and elliptic pseudodifferential operator (e.g. $C_U = (I - \Delta)^{-r}$). We assume that $U \sim N(0, C_U)$, that is we have a formal prior

$$\pi_{pr}(u) \underset{formally}{=} c \exp\left(-\frac{1}{2} \| C_U^{-1/2} u \|_{L^2}^2 \right) \tag{4}$$

i.e. we are interested of finding an approximation $u_\delta \in H^r$.

Above we assumed that the covariance operator $C_U \in \Psi^{-2r}$. Now the question is in what Sobolev space $H^{-\tau}$ does the prior U takes values?

Two definitions for covariance operator

if U takes values in $H^{-\tau}$ we can define the covariance operator of U two ways

1) $C_U : H^{\tau} \to H^{-\tau}$

$$\mathbb{E}\Big(\langle U, \phi\rangle_{H^{-\tau}\times H^{\tau}} \, \langle U, \psi\rangle_{H^{-\tau}\times H^{\tau}}\Big) = \langle C_U\phi, \psi\rangle_{H^{-\tau}\times H^{\tau}}$$

where $\langle\cdot,\cdot\rangle_{H^{-\tau}\times H^{\tau}}$ is a dual pairing.

2) $B_U : H^{-\tau} \to H^{-\tau}$

$$\mathbb{E}\big((U, \phi)_{H^{-\tau}} \, (U, \psi)_{H^{-\tau}}\big) = (B_U\phi, \psi)_{H^{-\tau}},$$

where $(\cdot,\cdot)_{H^{-\tau}}$ stands for the inner product.

The connection between B_U and C_U is

$$B_U = C_U(I - \Delta)^{-\tau} : H^{-\tau} \to H^{-\tau}.$$

The prior U takes values in $H^{-\tau}$, where $-\tau < r - d/2$

To guarantee that $U \in H^{-\tau}$ we will choose $-\tau \in \mathbb{R}$ so that

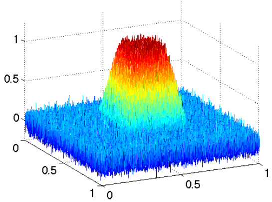
$$\mathbb{E}\big(\|U\|^2_{H^{-\tau}}\big) < \infty.$$

The above condition is equivalent with assumption that the covariance operator B_U is a trace class operator in $H^{-\tau}$. This is true if $-\tau < r - d/2$.

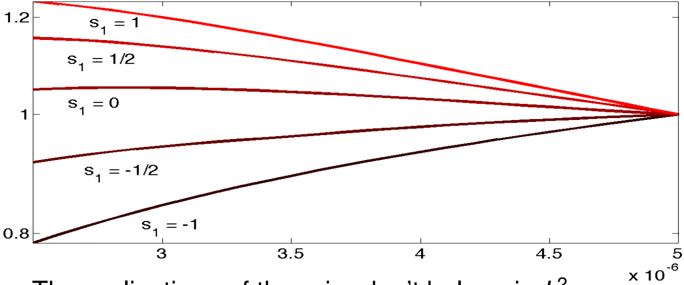
We have proven that when we are looking for an approximation $U_\delta \in H^r$ then the prior should take values in $H^{-\tau}$, where $-\tau < r - d/2$.

What does this mean for our example?

For Gaussian smoothness prior $r = 1$ but in two dimensional case we get that $-\tau < 0$.



An approximation $u_\delta \in H^1$



The realisations of the prior don't belong in L^2.

Theorem 1: Assumptions

We assume that

- N is a d-dimensional closed manifold.
- Operator C_U is $2r$ times smoothing, self-adjoint, injective and elliptic.
- Unknown U is a generalised random variable taking values in $H^{-\tau}$, $\tau > d/2 - r$ with mean zero and covariance operator C_U.
- $\mathcal{E}$ is white Gaussian noise taking values in H^{-s}, $s > d/2$.

Consider the measurement

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

where $A \in \Psi^{-t}$, is an elliptic pseudodifferential operator of order $-t < \min\{0, -\tau - s\}$. Assume that $A : L^2(N) \to L^2(N)$ is injective.

Theorem 1: Convergence results

Take $s_1 \leq r - s < r - d/2$. Then the following convergence takes place in $H^{s_1}(N)$ norm

$$\lim_{\delta \to 0} U_\delta(\omega) = U(\omega) \quad a.s.$$

Above $U_\delta(\omega) \in H^r$ a.s.

We have the following estimates for the speed of convergence:

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

$$\|U_\delta(\omega) - U(\omega)\|_{H^{s_1}} \leq C\delta \quad a.s.$$

(ii) If $-s - t \leq s_1 \leq r - s$ then

$$\|U_\delta(\omega) - U(\omega)\|_{H^{s_1}} \leq C\delta^{1 - \frac{s+t+s_1}{t+r}} \quad a.s.$$

Connection to Tikhonov regularisation

In deterministic framework one assumes that there exists a true solution $u \in H^r$. The regularised solution is given by

$$u_\delta = \arg\min_{u \in H^r} \left\{ \|Au\|_{L^2}^2 - 2\langle m, Au \rangle + \alpha \|u\|_{H^r}^2 \right\}$$

where $\alpha = \delta^\kappa$, $\kappa > 0$.

This corresponds to our previous MAP estimate if we choose $\kappa = 2$ and $C_U = (I - \Delta)^{-r}$.

Convergence results in Tikhonov case

Take $s_1 \leq \min\left\{r, \frac{2}{\kappa}r + \left(\frac{2}{\kappa} - 1\right)t - s\right\}$ then we get the following convergence

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

and the speed of convergence

$$\|u_\delta - u\|_{H^{s_1}} \leq C \max\{\delta^{\frac{\kappa(r-s_1)}{2(t+r)}}, \delta^{1-\frac{\kappa(s+t+s_1)}{2(t+r)}}\}.$$

Note that above $u, u_\delta \in H^r(N)$ with $r \geq 0$.

What can we say about the convergence depending on κ?

The regularised solution

$$u_\delta = \arg\min_{u \in H^r} \left\{ \|Au\|_{L^2}^2 - 2\langle m, Au \rangle + \delta^\kappa \|u\|_{H^r}^2 \right\}$$

can be written in a form

$$\begin{aligned} u_\delta &= K^{-1} A^* A u + K^{-1} A^* \delta \varepsilon \\ &= u_{noiseless} + u_{noise} \end{aligned}$$

where $K = A^* A + \delta^\kappa (I - \Delta)^r$.

- The noise term is dominating if we choose $\kappa \geq 2$.
- We get convergence in H^r when $\kappa \leq 1$.

Note that in classical theory we get convergence when $\kappa < 2$.

In a nutshell:
2-dimensional Gaussian smoothness prior

Inverse problem: $M = AU + \mathcal{E}\delta$, $\mathcal{E}$ white noise.

MAP estimate: Using a modified Tikhonov regularisation with penalty term $\|u\|_{H^1}$ we get MAP estimate $u_\delta^{MAP} \in H^1$.

Prior distribution: To get formally the above prior we assume that $U \sim N(0, C_U)$, where $C_U = (I - \Delta)^{-1}$.

CM estimate: The conditional distribution $u|m$ has mean $u_\delta^{CM} = u_\delta^{MAP} \in H^1$.

Space of the prior: When $C_U = (I - \Delta)^{-1}$ we can show that the prior U takes values in $H^{-\tau}$, $\tau > 0$.

Theorem 1: We can prove convergence speed in spaces H^{s_1} where $s_1 < 0$. Note that we don't necessarily get convergence in H^1 or even in L^2.