Bayesian Inverse Problems and Uncertainty Quantification

Hanne Kekkonen

Centre for Mathematical Sciences
University of Cambridge

June 4, 2019
Inverse problems arise naturally from applications
Inverse problems are ill-posed

We want to recover the unknown $u$ from a noisy measurement $m$;

$$m = Au + noise,$$

where $A$ is a forward operator that usually causes loss of information.
Inverse problems are ill-posed

We want to recover the unknown $u$ from a noisy measurement $m$:

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

where $A$ is a forward operator that usually causes loss of information.

Well-posedness as defined by Jacques Hadamard:

1. Existence: There exists at least one solution.
2. Uniqueness: There is at most one solution.
3. Stability: The solution depends continuously on data.

Inverse problems are **ill-posed** breaking at least one of the above conditions.
The naive inversion does not produce stable solutions

We want to approximate $u$ from a measurement

$$m = Au + n,$$

where $A : X \rightarrow Y$ is linear and $n$ is noise.

One approach is to use the least squares method

$$\tilde{u} = \arg \min_{u \in X} \{ ||Au - m||^2_Y \}.$$

**Problem:** Multiple minima and sensitive dependence on the data $m$. 
Tikhonov regularisation is a classical method for solving ill-posed problems

We want to approximate \( u \) from a measurement

\[
m = Au + n,
\]

where \( A : X \to Y \) is linear and \( n \) is noise.

The problem is ill-posed so we add a regularising term and get

\[
\tilde{u} = \arg \min_{u \in E \subset X} \left\{ \|Au - m\|_Y^2 + \alpha \|u\|_E^2 \right\}
\]

Regularisation gives a stable approximate solution for the inverse problem.
Bayes formula combines data and a priori information

We want to reconstruct the most probable $u \in \mathbb{R}^k$ in light of

- **Measurement information:** $M \mid u \sim P_u$ with Lebesgue density $\rho(m \mid u) = \rho_\varepsilon(m - Au)$.
- **A priori information:** $U \sim \Pi_{pr}$ with Lebesgue density $\pi_{pr}(u)$.

**Bayes’ formula**

We can update the prior, given a measurement, to a posterior distribution using the Bayes’ formula:

$$
\pi(u \mid m) \propto \pi_{pr}(u) \rho(m \mid u)
$$

The result of Bayesian inversion is the posterior distribution $\pi(u \mid m)$. 

The result of Bayesian inversion is the posterior distribution, but typically one looks at estimates

Maximum a posteriori (MAP) estimate:
\[ \arg \max_{u \in \mathbb{R}^n} \pi(u | m) \]

Conditional mean (CM) estimate:
\[ \int_{\mathbb{R}^n} u \pi(u | m) \, du \]
Gaussian example

Assume we are interested in the measurement model $M = AU + N$, where:

- $A : X \rightarrow Y$, with $X = \mathbb{R}^d$ and $Y = \mathbb{R}^k$.
- $N$ is white Gaussian noise.
- $U$ follows Gaussian prior.

Posterior has density

$$
\pi^m(u) = \pi(u \mid m) \propto \exp \left( - \frac{1}{2} \| m - Au \|^2_{\mathbb{R}^k} - \frac{1}{2} \| u \|^2_{\Sigma} \right)
$$

We can use the mean of the posterior as a point estimator but having the whole posterior allows uncertainty quantification.
Why are we interested in uncertainty quantification?
Uncertainty quantification has many applications

Studying the whole posterior distribution instead of just a point estimate offers us more information.

Uncertainty quantification
- Confidence and credible sets
- E.g. Weather and climate predictions

Using the whole posterior
- Geological sensing
- Bayesian search theory

Figure: Search for the wreckage of Air France flight AF 447, Stone et al.
What do we mean by uncertainty quantification?

-I’m going to die?
-POSSIBLY.
-Possibly? You turn up when people are possibly going to die?
-OH, YES. IT’S QUITE THE NEW THING. IT’S BECAUSE OF THE UNCERTAINTY PRINCIPLE.
-What’s that?
-I’M NOT SURE.
-That’s very helpful.
-I THINK IT MEANS PEOPLE MAY OR MAY NOT DIE. I HAVE TO SAY IT’S PLAYING HOB WITH MY SCHEDULE, BUT I TRY TO KEEP UP WITH MODERN THOUGHT.

-Terry Pratchett, *The Fifth Elephant*
A Bayesian credible set is a region in the posterior distribution that contains a large fraction of the posterior mass.
Frequentist confidence region
Once we have achieved a Bayesian solution the natural next step is to consider the consistency of the solution.

- **Convergence** of a point estimator to the ‘true’ $u^\dagger$.
- **Contraction** of the posterior distribution; Do we have

  $$\Pi(u : d(u, u^\dagger) > \delta_n | m) \rightarrow_{P_{u^\dagger}} 0,$$

  for some $\delta_n \rightarrow 0$, as the sample size $n \rightarrow \infty$.

Is optimal contraction rate enough to guarantee that the Bayesian credible sets have correct frequentist coverage?
Credible sets do not necessarily cover the truth well

Do credible sets quantify frequentist uncertainty?

Do we have for \( C = C(m) \)

\[
\Pi \left( u \in C \mid m \right) \approx 0.95 \quad \Leftrightarrow \quad P_{u^\dagger} \left( u^\dagger \in C(m^\dagger) \right) \approx 0.95?
\]
Do credible sets quantify frequentist uncertainty?

Do we have for \( C = C(m) \)

\[
\Pi\left( u \in C \mid m \right) \approx 0.95 \iff P_{u^\dagger}\left( u^\dagger \in C(m^\dagger) \right) \approx 0.95?
\]

**Bernstein–von Mises Theorem (BvM)**

For large sample size \( n \), with \( \hat{u}_{MLE} \) being the maximum likelihood estimator,

\[
\Pi(\cdot \mid m) \approx N\left( \hat{u}_{MLE}, \frac{1}{n}I(u^\dagger)^{-1} \right), \quad \text{for } M \sim P_{u^\dagger},
\]

whenever \( u^\dagger \in \mathcal{O} \subset \mathbb{R}^d \) and the prior \( \Pi \) has positive density on \( \mathcal{O} \), and the inverse Fisher information \( I(u^\dagger) \) is invertible.
BvM guarantees confident credible sets

The contraction rate of the posterior distribution near $u^\dagger$ is

$$\Pi \left( u : \|u - u^\dagger\|_{\mathbb{R}^d} \geq \frac{L_n}{\sqrt{n}} | m \right) \rightarrow P_{u^\dagger} 0 \quad \text{as } L_n, n \rightarrow \infty$$

For a fixed $d$ and large $n$ computing posterior probabilities is roughly the same as computing them from $N\left( \hat{u}_{MLE}, \frac{1}{n}I(f^\dagger)^{-1} \right)$
BvM guarantees confident credible sets

The contraction rate of the posterior distribution near $u^\dagger$ is

$$\Pi\left(u : \|u - u^\dagger\|_{\mathbb{R}^d} \geq \frac{L_n}{\sqrt{n}} | m\right) \rightarrow^{P_{u^\dagger}} 0 \quad \text{as } L_n, n \rightarrow \infty$$

For a fixed $d$ and large $n$ computing posterior probabilities is roughly the same as computing them from $N\left(\hat{u}_{MLE}, \frac{1}{n}I(f^\dagger)^{-1}\right)$

$$C_n \text{ s.t. } \Pi\left(u \in C_n | M\right) = 0.95 \quad \implies \quad P_{u^\dagger}\left(u^\dagger \in C_n\right) \rightarrow 0.95 \quad \text{(Bayesian credible set)}$$

$$P_{u^\dagger}\left(u^\dagger \in C_n\right) \rightarrow 0.95 \quad \text{(Frequentist confident set)}$$

$$|C_n|_{\mathbb{R}^d} = O_{P_{u^\dagger}}\left(\frac{1}{\sqrt{n}}\right) \quad \text{(Optimal diameter)}$$
Asymptotic normality of the Tikhonov regulariser

We return to the Gaussian example where the posterior is also Gaussian. The posterior mean $\bar{u}$ equals the MAP estimate which equals the Tikhonov-regulariser

$$\bar{u} = \arg\min_u \{ \|Au - m\|_{\mathbb{R}^k}^2 + \|u\|_\Sigma^2 \}.$$

Then the following convergence holds under $P_{u^\dagger}$

$$\sqrt{n}(\bar{u} - u^\dagger) \rightarrow Z \sim N(0, I(u^\dagger)^{-1})$$

as $n \rightarrow \infty$. 
Confident credible sets

We can now construct a confidence set for Tikhonov regulariser: Consider a credible set

\[ C_n = \left\{ u \in \mathbb{R}^d : \| u - \overline{u} \| \leq \frac{R_n}{\sqrt{n}} \right\}, \quad R_n \text{ s.t. } \Pi(C_n | m) = 0.95. \]

Then the frequentist coverage probability of \( C_n \) will satisfy

\[ P_{u^\dagger} \left( u^\dagger \in C_n \right) \to 0.95 \quad \text{and} \quad R_n \to P_{u^\dagger} \Phi^{-1}(0.95) \]

as \( n \to \infty \). Here \( \Phi^{-1} \) is a continuous inverse of \( \Phi = \mathbb{P}(Z \leq \cdot) \) with \( Z \sim N(0, I(u^\dagger)^{-1}) \).
Discretisation of $m$ is given by the measurement device but the discretisation of $u$ can be chosen freely.

$m \in \mathbb{R}^k$

$u \in \mathbb{R}^n$

$k = 4$

$n = 48$
The discretisations are independent

\[ m \in \mathbb{R}^k \]
\[ u \in \mathbb{R}^n \]

\[ k = 8 \]
\[ n = 156 \]
The discretisations are independent

\[ m \in \mathbb{R}^k \]
\[ u \in \mathbb{R}^n \]

\[ k = 24 \]
\[ n = 440 \]
The measurement is always discrete but the unknown is usually a continuous function

\[ m \in \mathbb{R}^4 \]
\[ u \in L^2 \]
We often want to use a continuous model for theory

\[ m = Au + \varepsilon \]
Nonparametric models

- In many applications it is natural to use a statistical regression model

\[ M_i = (AU)(x_i) + N_i, \quad i = 1, \ldots, n, \quad N_i \sim \mathcal{N}(0, 1), \]

where \( x_i \in \mathcal{O} \) are measurement points and \( A \) is a forward operator. The goal is to infer \( U \) from the data \( (M_i) \).
Nonparametric models

- In many applications it is natural to use a statistical regression model

\[ M_i = (AU)(x_i) + N_i, \quad i = 1, \ldots, n, \; N_i \sim \mathcal{N}(0, 1), \]

where \( x_i \in \mathcal{O} \) are measurement points and \( A \) is a forward operator. The goal is to infer \( U \) from the data \((M_i)\).

- For the theory we use a continuous model, which corresponds \((x_i)\) growing dense in the domain \( \mathcal{O} \).

- If \( \mathbb{W} \) is a Gaussian white noise process in the Hilbert space \( \mathcal{H} \) then

\[ M = AU + \varepsilon \mathbb{W}, \quad \varepsilon = \frac{1}{\sqrt{n}} \text{ noise level}, \quad M \sim P_{u^\dagger} \]

Note that usually \( Au \in L^2 \) but \( \mathbb{W} \in H^{-s} \) only with \( s > d/2 \).
Gaussian priors are often used for inverse problems

- Gaussian priors $\Pi$ are often used in practice: see e.g Kaipio & Somersalo (2005), Stuart (2010), Dashti & Stuart (2016).
Gaussian priors are often used for inverse problems

- Gaussian priors $\Pi$ are often used in practice: see e.g Kaipio & Somersalo (2005), Stuart (2010), Dashti & Stuart (2016).

- Using the Cameron-Martin theorem we can formally write

  $$ d\Pi(\cdot | m) \propto e^{\ell(u)} d\Pi(u) \propto e^{\ell(u) - \frac{1}{2}\|u\|_V^2}, $$

  where $\ell(u) = \frac{1}{\varepsilon^2} \langle m, Au \rangle - \frac{1}{2\varepsilon^2} \|Au\|^2$, and $V_\Pi$ denotes the Cameron-Martin space of $\Pi$.
Gaussian priors are often used for inverse problems

- Gaussian priors $\Pi$ are often used in practice: see e.g Kaipio & Somersalo (2005), Stuart (2010), Dashti & Stuart (2016).

- Using the Cameron-Martin theorem we can formally write

$$d\Pi(\cdot \mid m) \propto e^{\ell(u)} d\Pi(u) \propto e^{\ell(u) - \frac{1}{2} \|u\|_{V_{\Pi}}^2},$$

where $\ell(u) = \frac{1}{\varepsilon^2} \langle m, Au \rangle - \frac{1}{2\varepsilon^2} \|Au\|^2$, and $V_{\Pi}$ denotes the Cameron-Martin space of $\Pi$.

- The Cameron-Martin space characterises the directions in which a Gaussian measure can be shifted to obtain an equivalent Gaussian measure.

- If $U \sim \mathcal{N}(0, \Sigma)$ then $\|u\|_{V_{\Pi}}^2 = \|\Sigma^{-1/2}u\|_{L^2}^2$. 

If \( u \) is a function the classical BvM theorem does not hold

**Semi-parametric approach**

For fixed test function \( \psi \in \Psi \) we study the induced posterior distribution for the one-dimensional variable

\[
\langle U, \psi \rangle, \quad U \sim \Pi(\cdot \mid m).
\]

The idea is to determine possibly maximal families \( \Psi \) of functions \( \psi \) for which the Gaussian asymptotics

\[
\frac{1}{\varepsilon} \left( \langle U, \psi \rangle - \hat{u}(m) \mid m \right) \to Z \sim N(0, I(u^\dagger, \psi)^{-1})
\]

can be obtained, as \( \varepsilon \to 0 \). Above \( \hat{u}(m) \) is an efficient estimator of \( \langle u^\dagger, \psi \rangle \).

This approach has been used in recent papers Castillo & Nickl (2013, 2014), Monard, Nickl & Paternain (2019), Nickl (2018) and Giordano & Kekkonen (2018)
Example: elliptic boundary value problem

Let $\mathcal{O} \subset \mathbb{R}^d$ be bounded with $C^\infty$ boundary $\partial \mathcal{O}$. We are interested in recovering the unknown source $f \in L^2(\mathcal{O})$ in

$$\begin{cases}
  Lv = -\nabla \cdot (\sigma \nabla v) = f & \text{on } \mathcal{O} \\
  v = 0 & \text{on } \partial \mathcal{O}
\end{cases}$$

from noisy observations $M = L^{-1}f + \varepsilon W$. 
Example: elliptic boundary value problem

Let $\mathcal{O} \subset \mathbb{R}^d$ be bounded with $C^\infty$ boundary $\partial \mathcal{O}$. We are interested in recovering the unknown source $f \in L^2(\mathcal{O})$ in

$$\begin{cases} L v = -\nabla \cdot (\sigma \nabla v) = f & \text{on } \mathcal{O} \\ v = 0 & \text{on } \partial \mathcal{O} \end{cases}$$

from noisy observations $M = L^{-1}f + \varepsilon W$.

- The forward operator is $A = L^{-1} : L^2 \rightarrow L^2$ is smoothing of order 2.

- We assign $f$ a "correctly specified" centred Gaussian prior $\Pi$ on $L^2$. That is, $f^\dagger \in H^\alpha$, $\alpha \geq 0$, and $\Pi$ has Cameron-Martin space $V_\Pi = H^r$ with $\frac{d}{2} \leq r \leq \alpha + \frac{d}{2}$. 
Let $\Pi$ be the Gaussian prior described above and $\Pi(\cdot \mid M)$ the resulting posterior distribution.

**Theorem 1 (Giordano and K. 2018)**

If $f \sim \Pi(\cdot \mid M)$ and $\bar{f} = \mathbb{E}_{\Pi}(f \mid M)$, then for all $\psi \in H_c^\beta$, $\beta > 2 + d/2$,

$$
\mathcal{L} \left( \frac{1}{\varepsilon} \langle f - \bar{f}, \psi \rangle_{L^2} \mid M \right) \overset{\mathcal{L}}{\rightarrow} N(0, \|L\psi\|_{L^2}^2)
$$

in $P_{f^\dagger}^M$-probability as $\varepsilon \to 0$.

Here we have denoted $\psi \in H_c^\beta = \{\psi \in H^\beta, \text{supp}(\psi) \subsetneq \mathcal{O}\}$. 