Bayesian inverse problems with Laplacian noise

Remo Kretschmann

Faculty of Mathematics, University of Duisburg-Essen

Applied Inverse Problems 2017, M27
Hangzhou, 1 June 2017

Outline

Classical setting

- X, Y separable Hilbert spaces,
- $F\colon X \to Y$.

Given **observed data** $y \in Y$ find **unknown** $u \in X$, where

$$y = F(u) + \eta$$

with **observational noise** $\eta \in Y$.

Bayesian approach

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

- X, Y separable Hilbert spaces,
- $F\colon X \to Y$,
- probability measures μ_0 on $(X, \mathcal{B}(X))$, $\mathbb{Q}_0$ on $(Y, \mathcal{B}(Y))$,
- **prior** $u \sim \mu_0$, **noise** $\eta \sim \mathbb{Q}_0$, η independent of u and

$$y = F(u) + \eta.$$

Given **observed data** $y \in Y$ find **posterior distribution** μ^y, the conditional distribution of $u|y$.

- Extract information out of μ^y in the form of estimators.

Motivation

Bayesian inverse problems in function spaces:

- Dashti, Law, Stuart and Voss have studied nonlinear inverse problems with Gaussian prior and noise that satisfies certain conditions. [Dashti et al 2013]
- In this case, the MAP estimator can be described as the minimiser of the Onsager-Machlup functional.
- Dashti and Stuart have analysed the inverse heat equation with Gaussian noise and different priors (i.a. Gaussian). [Dashti, Stuart 2015]

Motivation

Bayesian inverse problems in function spaces:

- Dashti, Law, Stuart and Voss have studied nonlinear inverse problems with Gaussian prior and noise that satisfies certain conditions. [Dashti et al 2013]
- In this case, the MAP estimator can be described as the minimiser of the Onsager-Machlup functional.
- Dashti and Stuart have analysed the inverse heat equation with Gaussian noise and different priors (i.a. Gaussian). [Dashti, Stuart 2015]

Questions:

- What happens if the prior is Gaussian but the noise is non-Gaussian?
- Does Laplacian noise lead to an ℓ^1-discrepancy term?

Motivation

- We study the **inverse heat equation** with **Laplacian noise** in combination with a **Gaussian prior**.
- **Problem:** Laplacian noise violates the conditions of [Dashti et al 2013].
- Existence of a solution?
- Connection: MAP estimator – optimisation problem?
- Does the MAP estimator converge towards the true solution, as the variance of the noise tends to zero?

The heat conduction equation

- $D \subset \mathbb{R}^d$ bounded domain, $\partial D \in C^k$ for some $k \geqslant 1$,
- $A := -\Delta$ defined on $\mathcal{D}(A) = H^2(D) \cap H_0^1(D)$.

For every $u \in L^2(D)$ there is a unique solution

$$v \in C([0,\infty), L^2(D)) \cap C^1((0,\infty), \mathcal{D}(A))$$

of the heat equation on D with Dirichlet boundary conditions,

$$\begin{cases} \dfrac{\mathrm{d}v}{\mathrm{d}t}(t) = -Av(t) & \text{for } t > 0, \\ v(0) = u, & \end{cases}$$

given by

$$v(t) = \exp(-At)u \qquad \text{for all } t \geqslant 0.$$

- Fix $t = 1$, i.e. $F(u) = v(1) = e^{-A}u$.

Given temperature measurement y at time $t = 1$, find initial temperature $u \in L^2(D)$ at time $t = 0$, where

$$y = e^{-A}u + \eta.$$

The Bayesian inverse problem (outline)

Given temperature measurement y at time $t = 1$, find conditional distribution of the **posterior** $u|y$, where

$$y = e^{-A}u + \eta.$$

We assume that

- $-A$ is a Laplace-like operator,
- the **noise** η has a **centred Laplacian distribution** with covariance operator $A^{s-\beta}$, and
- the **prior** u has a **centred Gaussian distribution** with covariance operator $A^{-\tau}$.

Laplace-like operators

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

We assume that the operator A in $L^2(D)$ satisfies the following properties:

1. The eigenvectors $\{\varphi_k\}_{k\in\mathbb{N}}$ of A form an orthonormal basis of $L^2(D)$.
2. The respective eigenvalues $\alpha_1 \geqslant \alpha_2 \geqslant \cdots > 0$ of A satisfy

$$\frac{1}{C_A} k^{\frac{2}{d}} \leqslant \alpha_k \leqslant C_A k^{\frac{2}{d}} \qquad \text{for all } k \in \mathbb{N}$$

   and a constant $C_A > 1$.
3. A is densely defined and surjective.
4. A is self-adjoint.

Hilbert scales

A induces a **Hilbert scale** $\{\mathcal{H}^s\}_{s\in\mathbb{R}}$, where

$$\mathcal{H}^s := A^{-s}(L^2(D)) = \left\{u \in L^2(D) : \sum_{k=1}^{\infty} \alpha_k^{2s} \left|(u, \varphi_k)_{L^2}\right|^2 < \infty\right\}$$

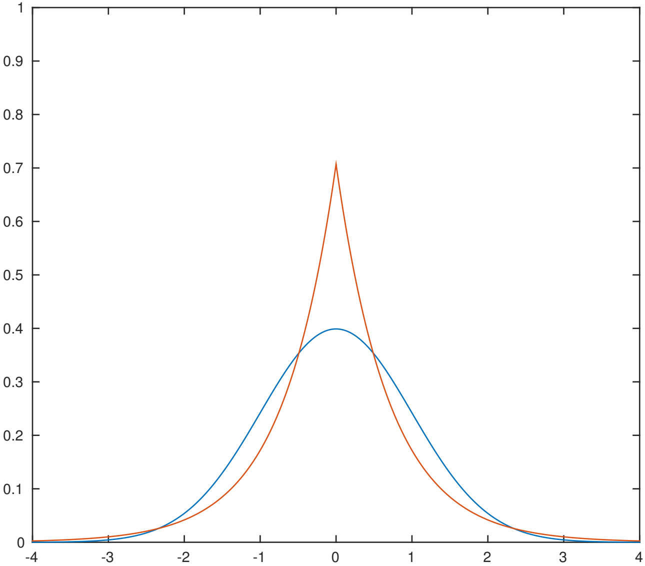
for all $s \geqslant 0$, equipped with

$$\|u\|_{\mathcal{H}^s} := \|A^{\frac{s}{2}} u\|_{L^2} \quad \text{and} \quad (u, v)_{\mathcal{H}^s} := (A^{\frac{s}{2}} u, A^{\frac{s}{2}} v)_{L^2}.$$

Now we set $X := L^2(D) = \mathcal{H}^0$ and $Y := \mathcal{H}^s$ with $s \geqslant 0$, i.e.,

$$u \in L^2(D) \quad \text{and} \quad \eta, y \in \mathcal{H}^s.$$

Standard Laplacian measure on $\mathbb{R}$



Laplacian measure on $\mathbb{R}$

For $a \in \mathbb{R}$ and $\lambda > 0$ define probability measure $\mathcal{L}_{a,\lambda}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by

$$\mathcal{L}_{a,\lambda}(B) = \frac{1}{\sqrt{2\lambda}} \int_B e^{-\frac{\sqrt{2}|x-a|}{\sqrt{\lambda}}} \mathrm{d}x \qquad \text{for all } B \in \mathcal{B}(\mathbb{R}).$$

Then $\mathcal{L}_{a,\lambda}$ has **mean** a and **variance** λ, i.e.,

$$\int_{\mathbb{R}} x \mathcal{L}_{a,\lambda}(\mathrm{d}x) = a, \qquad \int_{\mathbb{R}} (x-a)^2 \mathcal{L}_{a,\lambda}(\mathrm{d}x) = \lambda.$$

Infinite-dimensional product measure

- H separable real Hilbert space.
- For every compact self-adjoint operator Q on H there is an orthonormal basis $\{e_k\}_{k\in\mathbb{N}}$ of H consisting of eigenvectors of Q.
- Identify H with ℓ^2 by means of $x \mapsto \{(x, e_k)_H\}_{k\in\mathbb{N}}$.

Idea: For any $a \in H$ and any positive definite trace class operator $Q \in \mathcal{L}(H)$ define **Laplacian measure** $\mathcal{L}_{a,Q}$ on $(\ell^2, \mathcal{B}(\ell^2))$ as the product measure

$$\mathcal{L}_{a,Q} = \bigotimes_{k=1}^{\infty} \mathcal{L}_{a_k,\lambda_k},$$

with $a_k := (a, e_k)_H$ and $\lambda_k := (Qe_k, e_k)_H$ for all $k \in \mathbb{N}$.

Infinite-dimensional product measure

- H separable real Hilbert space.
- For every compact self-adjoint operator Q on H there is an orthonormal basis $\{e_k\}_{k\in\mathbb{N}}$ of H consisting of eigenvectors of Q.
- Identify H with ℓ^2 by means of $x \mapsto \{(x, e_k)_H\}_{k\in\mathbb{N}}$.

Idea: For any $a \in H$ and any positive definite trace class operator $Q \in \mathcal{L}(H)$ define **Laplacian measure** $\mathcal{L}_{a,Q}$ on $(\ell^2, \mathcal{B}(\ell^2))$ as the product measure

$$\mathcal{L}_{a,Q} = \bigotimes_{k=1}^{\infty} \mathcal{L}_{a_k,\lambda_k},$$

with $a_k := (a, e_k)_H$ and $\lambda_k := (Qe_k, e_k)_H$ for all $k \in \mathbb{N}$.

Caution: This definition depends on the choice of the basis $\{e_k\}_{k\in\mathbb{N}}$.

Basic properties

This way, $\mathcal{L}_{a,Q}$ has **mean** a and **covariance operator** Q, i.e.,

$$\int_H (x,y)_H \mathcal{L}_{a,Q}(\mathrm{d}x) = (a,y)_H \qquad \text{for all } y \in H,$$

$$\int_H (x-a,y)_H (x-a,z)_H \mathcal{L}_{a,Q}(\mathrm{d}x) = (Qy,z)_H \qquad \text{for all } y,z \in H.$$

In case $a = 0$ we write $\mathcal{L}_Q := \mathcal{L}_{0,Q}$.

The Bayesian inverse problem

Given $y \in Y$, find conditional distribution of $u|y$ on X, where

- **noise** $\eta \sim \mathcal{L}_{A^{s-\beta}}$ with **Laplacian measure** $\mathcal{L}_{A^{s-\beta}}$ on $Y := \mathcal{H}^s$ using basis $e_k := \alpha_k^{-\frac{s}{2}} \varphi_k$ and $0 \leqslant s < \beta - \frac{d}{2}$,
- **prior** $u \sim \mathcal{N}_{A^{-\tau}}$ independent from η with **Gaussian measure** $\mathcal{N}_{A^{-\tau}}$ on $X := L^2(D) = \mathcal{H}^0$ and $\tau > \frac{d}{2}$,
- $y = e^{-A} u + \eta$.

Idea: Use Bayes' Theorem to obtain posterior distribution.

Outline

Bayes' Theorem

- (X, A), (Y, B) measurable spaces,
- ν, ν_0 probability measures on $X \times Y$, such that $\nu \ll \nu_0$, i.e., ν is **absolutely continuous with respect to** ν_0.
- Then ν has a density $f = \frac{\mathrm{d}\nu}{\mathrm{d}\nu_0}$ with respect to ν_0, i.e., $\nu = f\nu_0$.

Theorem (Bayes)

Assume that the conditional random variable $x|y$ exists under ν_0 with probability distribution ν_0^y on X. Then the conditional random variable $x|y$ exists under ν with probability distribution ν^y on X, and $\nu^y \ll \nu_0^y$. If additionally, $Z(y) := \int_X \frac{\mathrm{d}\nu}{\mathrm{d}\nu_0}(x, y)\nu_0^y(\mathrm{d}x) > 0$, then

$$\frac{\mathrm{d}\nu^y}{\mathrm{d}\nu_0^y}(x) = \frac{1}{Z(y)} \frac{\mathrm{d}\nu}{\mathrm{d}\nu_0}(x, y).$$

Posterior distribution

- In our case, $(u, \eta) \sim \nu_0$ and $(u, y) \sim \nu$ on $X \times Y = L^2(D) \times \mathcal{H}^s$.
- In order for $\nu \ll \nu_0$ to hold, we require $\mathcal{L}_{e^{-A}u, A^{s-\beta}} \ll \mathcal{L}_{A^{s-\beta}}$ for all $u \in X$.
- Then by Bayes' Theorem, the **posterior measure** μ^y of $u|y$ is absolutely continuous with respect to the prior measure $\mathcal{N}_{A^{-\tau}}$ with the density

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mathcal{N}_{A^{-\tau}}}(u) = \frac{1}{Z(y)} \exp(-\Phi(u, y)) \qquad \nu_0\text{-a.e.},$$

$$\Phi(u, y) = \sqrt{2} \sum_{k=1}^{\infty} \alpha_k^{\frac{\beta}{2}} \left(|y_k - e^{-\alpha_k} u_k| - |y_k|\right),$$

where $y_k := (y, \varphi_k)_X$, $u_k := (u, \varphi_k)_X$.

Admissible shifts

- H separable Hilbert space, $Q \in \mathcal{L}(H)$ positive definite trace class operator.

Theorem

1. If $a \notin Q^{\frac{1}{2}}(H)$ then $\mathcal{L}_{a,Q}$ and $\mathcal{L}_Q$ are singular.
2. If $a \in Q^{\frac{1}{2}}(H)$ then $\mathcal{L}_{a,Q}$ and $\mathcal{L}_Q$ are equivalent ($\mathcal{L}_{a,Q} \ll \mathcal{L}_Q$ and $\mathcal{L}_Q \ll \mathcal{L}_{a,Q}$) and

$$\frac{d\mathcal{L}_{a,Q}}{d\mathcal{L}_Q}(y) = \exp\left(-\sqrt{2}\sum_{k=1}^{\infty}\frac{|y_k - a_k| - |y_k|}{\sqrt{\lambda_k}}\right) \qquad \mathcal{L}_Q\text{-a.e.},$$

where $y_k := (y, e_k)_H$, $a_k := (a, e_k)_H$ and $\lambda_k = (Qe_k, e_k)_H$.

Idea of proof: Apply Kakutani's Theorem.

Outline

Maximum a posteriori estimators

Let μ be a probability measure on a separable Hilbert space X and define

$$M_\varepsilon := \sup_{u \in X} \mu(B_\varepsilon(u)) \qquad \text{for all } \varepsilon > 0.$$

Any point $\widehat{u} \in X$ satisfying

$$\lim_{\varepsilon \to 0} \frac{\mu(B_\varepsilon(\widehat{u}))}{M_\varepsilon} = 1$$

is called a **maximum a posteriori estimator** for μ.

Onsager-Machlup functional

$I\colon E \to \mathbb{R}$ is called **Onsager-Machlup functional** for μ, if

$$\lim_{\varepsilon \to 0} \frac{\mu(B_\varepsilon(u))}{\mu(B_\varepsilon(v))} = \exp(I(v) - I(u))$$

for all $u, v \in E$, where $E \subseteq X$ denotes the space of all admissible shifts that yield an equivalent measure.

Onsager-Machlup functional

$I\colon E \to \mathbb{R}$ is called **Onsager-Machlup functional** for μ, if

$$\lim_{\varepsilon\to 0} \frac{\mu(B_\varepsilon(u))}{\mu(B_\varepsilon(v))} = \exp(I(v) - I(u))$$

for all $u, v \in E$, where $E \subseteq X$ denotes the space of all admissible shifts that yield an equivalent measure.

- For a centered Gaussian measure $\mathcal{N}_Q$ on X, $E = Q^{\frac{1}{2}}(X)$ and

$$I(u) = \frac{1}{2}\|Q^{-\frac{1}{2}}u\|_X^2 \qquad \text{for all } u \in E$$

Characterisation of MAP estimators

- μ_0 **centred Gaussian measure** on X, μ^y posterior measure on X with $\frac{d\mu^y}{d\mu_0}(u) = \exp(-\Phi(u))$ μ_0-a.e., $\Phi\colon X \to \mathbb{R}$,
- X separable Banach space, $E \subseteq X$ space of admissible shifts for μ_0, μ^y that yield an equivalent measure.

Theorem [Dashti et al 2013]

Assume that

1. Φ is bounded from below,
2. Φ is locally bounded from above,
3. Φ is locally Lipschitz continuous.

Then $u \in E$ is a MAP estimator for μ^y if and only if it minimises the Onsager-Machlup functional I for μ^y.

Onsager-Machlup functional (2)

- For $\mu_0 = \mathcal{N}_{A^{-\tau}}$ and μ^y, the space of admissible shifts is given by

$$E = A^{-\frac{\tau}{2}}(L^2(D)) = \mathcal{H}^\tau.$$

- In our case, Onsager-Machlup functional $I \colon \mathcal{H}^\tau \to \mathbb{R}$ for μ^y,

$$\begin{aligned} I(u) &:= \Phi(u) + \frac{1}{2}\|u\|^2_{\mathcal{H}^\tau} \\ &= \sqrt{2}\sum_{k=1}^{\infty} \alpha_k^{\frac{\beta}{2}} \left(|y_k - e^{-\alpha_k} u_k| - |y_k|\right) + \frac{1}{2}\sum_{k=1}^{\infty} \alpha_k^{\tau} |u_k|^2, \end{aligned}$$

where $u_k := (u, \varphi_k)_{L^2}$ and $y_k := (y, \varphi_k)_{L^2}$.

Characterisation of MAP estimators (2)

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

Theorem [Dashti et al 2013]

Assume that

1. Φ is bounded from below,
2. Φ is locally bounded from above,
3. Φ is locally Lipschitz continuous.

Then $u \in E$ is a MAP estimator for μ^y if and only if it minimises the Onsager-Machlup functional I for μ^y.

Problem: For Laplacian noise, Φ is **not** bounded from below.

Upside: Φ is **globally** Lipschitz continuous.

Characterisation of MAP estimators (3)

- μ_0 **centred Gaussian measure** on X, μ^y posterior measure on X with $\frac{d\mu^y}{d\mu_0}(u) = \exp(-\Phi(u))$ μ_0-a.e., $\Phi\colon X \to \mathbb{R}$,
- X separable Hilbert space, $E \subseteq X$ space of admissible shifts for μ_0, μ^y that yield an equivalent measure.

Theorem

Assume that

1. Φ is globally Lipschitz continuous,
2. $\Phi(0) = 0$.

Then $u \in E$ is a MAP estimator for μ^y if and only if it minimises the Onsager-Machlup functional I for μ^y.

Characterisation of MAP estimators (4)

Idea of proof:

- Show that $\{u_\varepsilon\}_{\varepsilon>0}$,

$$u_\varepsilon := \arg\max_{u \in X} \mu^y(B_\varepsilon(u)),$$

  contains a subsequence $\{u_{\varepsilon_n}\}_{n\in\mathbb{N}}$ that converges in X and its limit $u_0 \in E$ is both a MAP estimator for μ^y and a minimiser of I.
- Show that every MAP estimator $\widehat{u} \in X$ also minimises I.
- Show that every minimiser $\bar{u} \in E$ of I also is a MAP estimator.

Consistency of the MAP estimator

- Does the MAP estimator converge towards the true solution, as the variance of the noise tends to zero?
- How to choose the variance of the prior appropriately?

Scaled distributions

- Noise distribution $\mathcal{L}_{b^2 A^{s-\beta}}$, $b > 0$
- prior distribution $\mathcal{N}_{r^2 A^{-\tau}}$, $r > 0$.
- Associated Onsager-Machlup functional $I\colon \mathcal{H}^\tau \to \mathbb{R}$ for $y \in \mathcal{H}^s$,

$$I(u) = \frac{1}{b}\Phi(u) + \frac{1}{2r^2}\|u\|^2_{\mathcal{H}^\tau}.$$

Scaled distributions

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

- Noise distribution $\mathcal{L}_{b^2 A^{s-\beta}}$, $b > 0$
- prior distribution $\mathcal{N}_{r^2 A^{-\tau}}$, $r > 0$.
- Associated Onsager-Machlup functional $I \colon \mathcal{H}^\tau \to \mathrm{I\!R}$ for $y \in \mathcal{H}^s$,

$$I(u) = \frac{1}{b}\Phi(u) + \frac{1}{2r^2}\|u\|^2_{\mathcal{H}^\tau}.$$

- By the previous theorem, every minimiser $\bar{u}(y)$ of I is a MAP estimator for μ^y.
- Its components are

$$(\bar{u}(y), \varphi_k)_{L^2} = \max\left\{-\frac{r^2}{b}c_k, \min\left\{e^{\alpha_k} y_k, \frac{r^2}{b}c_k\right\}\right\},$$

where $y_k = (y, \varphi_k)_{L^2}$, $c_k := \sqrt{2}\alpha_k^{\frac{\beta}{2}-\tau} e^{-\alpha_k}$.

Frequentist consistency

- True solution $u^\dagger \in L^2(D)$ (fixed, no prior),
- positive sequences $\{b_n\}_{n\in\mathbb{N}}$, $\{r_n\}_{n\in\mathbb{N}}$ with $b_n \to 0$,
- Laplacian noise $\eta^n \in \mathcal{H}^s$ with $\eta^n \sim \mathcal{L}_{b_n^2 A^{s-\beta}}$ and

$$y^n = e^{-A}u^\dagger + \eta^n.$$

- Let u^n denote the respective minimisers of $I_n \colon \mathcal{H}^\tau \to \mathbb{R}$,

$$I_n(u) := \frac{1}{b_n}\Phi(u, y^n) + \frac{1}{2r_n^2}\|u\|^2_{\mathcal{H}^\tau}.$$

Convergence in mean square

Theorem

If a $w \in \mathcal{H}^{2\tau-\beta} \cap L^2(D)$ with $\|w\|_{\mathcal{H}^{2\tau-\beta}} \leqslant \rho$ exists, such that

$$u^\dagger = e^{-A} w,$$

and if $C > 0$ and $N \in \mathbb{N}$ exist, such that

$$\rho^{\frac{1}{2}} b_n^{\frac{1}{2}} \leqslant r_n \leqslant C^{\frac{1}{2}} b_n^{\frac{1}{2}} \qquad \text{for all } n \geqslant N,$$

then

$$\mathbb{E}\left[\|u^n - u^\dagger\|_{L^2}^2\right] \leqslant 2C \operatorname{Tr} A^{-\tau} b_n \qquad \text{for all } n \geqslant N.$$

Conclusion

Bayesian inverse heat equation with Laplacian noise:

- The posterior distribution exists.
- Every minimiser of the Onsager-Machlup functional is a MAP estimator.
- The MAP estimator is consistent in a frequentist sense.

Outlook:

- Conditional mean estimator in explicit form
- Direct posterior sampling