Generalised Modes in Bayesian Inverse Problems

Christian Clason[1] Tapio Helin[2] **Remo Kretschmann**[2]
Petteri Piiroinen[3]

[1]Faculty of Mathematics, University of Duisburg-Essen

[2]School of Engineering Science, Lappeenranta-Lahti University of Technology

[3]Department of Mathematics and Statistics, University of Helsinki

Inverse Days
Jyväskylä, 16. December 2019

Focus

- Maximum a posteriori (MAP) estimators in Bayesian inverse problems
- Nonparametric setting: Unknown quantity is modelled as element of infinite-dimensional space.

Bayesian inverse problems

- Reconstruct **unknown quantity** x from indirect **noisy measurement** y, where

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

- Probability distribution is assigned to unknown quantity (**Prior distribution**).
- Conditional distribution of unknown quantity, given measured data, is inferred using Bayes's formula (**Posterior distribution**).
- Obtain point estimates from this distribution, e.g., **MAP estimates** (modes of posterior distribution).

Nonparametric MAP estimates

- Define MAP estimate as **infinite-dimensional object** with certain **optimality properties** that describe its relation to posterior distribution.
- MAP estimate can be **approximated numerically**, while optimality properties are **independent of chosen discretisation**.
- Defining and establishing those optimality properties rigorously is **more challenging** in infinite-dimensional spaces.
- There are cases where **common approach fails**.

Bayesian inference in finite-dimensional spaces

- If **prior distribution** has density $x \mapsto \exp(-R(x))$ w.r.t. Lebesgue measure and
- **conditional distribution of** y **given** x has density $y \mapsto \exp(-\Phi(x; y))$,
- then **conditional distribution of** x **given** y, according to **Bayes's formula**, has density

$$x \mapsto \frac{1}{Z} \exp(-\Phi(x; y)) \exp(-R(x)).$$

Example

For **inverse problems** with **white Gaussian noise** we have

$$\Phi(x; y) = \frac{1}{2}\|F(x) - y\|^2.$$

Modes

- **Modes** of probability measure on **finite-dimensional space** are typically defined as **maximisers** of its **density** w.r.t. **Lebesgue measure**.
- **Modes** of **posterior distribution** are given as **minimisers** of

$$x \mapsto \Phi(x; y) + R(x).$$

Problem: There exists no Lebesgue measure on infinite-dimensional separable Banach spaces.

- For this reason, **modes** in **infinite-dimensional spaces** are typically defined via **asymptotic small ball probabilities** [Dashti, Law, Stuart, and Voss 2013; Lie and Sullivan 2018].

Modes in infinite-dimensional spaces

Let μ be a Borel probability measure on a separable Banach space X.

Definition

A point $\hat{x} \in X$ is called a **(strong) mode** of μ if

$$\lim_{\delta \to 0} \frac{\mu(B^\delta(\hat{x}))}{\sup_{x \in X} \mu(B^\delta(x))} = 1,$$

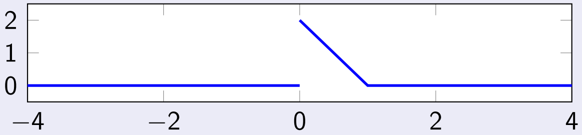
where $B^\delta(x)$ denotes the open ball around $x \in X$ with radius δ.

Problems with definition of modes

Example (measure without mode)

The probability measure μ on $\mathbb{R}$ with Lebesgue density

$$p(x) = \begin{cases} 2(1-x), & \text{if } x \in [0,1], \\ 0, & \text{otherwise.} \end{cases}$$



does not have a mode in 0, since

$$\lim_{\delta \to 0} \frac{\mu(B^\delta(0))}{\sup_{x \in \mathbb{R}} \mu(B^\delta(x))} = \lim_{\delta \to 0} \frac{\mu(B^\delta(0))}{\mu(B^\delta(\delta))} = \frac{1}{2}.$$

Practical relevance

There are applications where **strict bounds** on the admissible values of the unknown x emerge in a natural way, e.g.,

- radiography,
- electrical impedance tomography.

Such bounds can translate into **discontinuities** of the **posterior distribution**.

Bayesian inference in infinite-dimensional spaces

- **Prior distribution** μ_0 on separable Banach space X.
- If there exists **reference measure** ν (e.g., noise distribution) such that for all $x \in X$, **conditional distribution of** y **given** x has density

$$y \mapsto \exp(-\Phi(x; y))$$

  w.r.t. ν,
- then **conditional distribution of** x **given** y has the form

$$\mu^y(\mathrm{d}x) = \frac{1}{Z} \exp(-\Phi(x; y)) \mu_0(\mathrm{d}x)$$

  according to **Bayes's formula**.

Variational characterisation of MAP estimates

Under certain conditions, **modes** of **posterior distribution** are precisely **minimisers** of functional

$$I(x) = \Phi(x; y) + R(x)$$

in case of **Gaussian prior** [Dashti, Law, Stuart, and Voss 2013; Kretschmann 2019] or **Besov prior** [Agapiou, Burger, Dashti, and Helin 2018].

Goals

1. **Extend definition** of **modes** and thereby of **MAP estimates in nonparametric Bayesian inference** to cover cases where previous approach fails.
2. Show that our definition **coincides** with the previous one for a number of **commonly used prior measures** and find general **conditions** for coincidence.
3. Show that **generalised MAP estimates** are given as **minimisers** of canonical objective functional for **posterior distributions** with **discontinuities**.
4. Show **consistency** of generalised MAP estimator for **Bayesian inverse problems**.

Generalised modes

Idea: Replace fixed center point $\hat{x}$ in definition of mode by **sequence of center points** $\{w_\delta\}_{\delta>0}$ that converges to $\hat{x}$ as $\delta \to 0$.

Let μ be a Borel probability measure on a separable Banach space X.

Definition

A point $\hat{x} \in X$ is called a **generalised mode** of μ if for every positive sequence $\{\delta_n\}_{n\in\mathbb{N}}$ with $\delta_n \to 0$ there exists a sequence $\{w_n\}_{n\in\mathbb{N}} \subset X$ such that $w_n \to \hat{x}$ in X and

$$\lim_{n\to\infty} \frac{\mu(B^{\delta_n}(w_n))}{\sup_{x\in X} \mu(B^{\delta_n}(x))} = 1.$$

We call such a sequence $\{w_n\}_{n\in\mathbb{N}}$ an **approximating sequence**.

Generalised modes

Examples

- In the previous example, $\hat{x} := 0$ is a generalised mode with $w_n := \delta_n$.
- For Gaussian measures, the strong mode is the only generalised mode.

Criteria for coincidence of strong and generalised modes

When is a **generalised mode** $\hat{x}$ a **strong mode**?

1. Fundamental criterion.
2. Criterion using **convergence rate** of approximating sequence $\{w_n\}$.
3. Criterion using **convergence** of approximating sequence $\{w_n\}$ **in subspace topology**.

Variational characterisation of GMAP estimates

- **Goal:** Characterise **generalised modes** of **posterior distributions** which display **discontinuities** as **minimisers** of appropriate objective functional.
- Specifically, consider **prior** in separable subspace of ℓ^∞ with **strictly bounded** and **uniformly distributed components**.
- For this **prior distribution**, strong and generalised modes **do not coincide**.

Prior distribution

- Let $\{\gamma_k\}_{k\in\mathbb{N}}$ be a **sequence of weights** with

$$\gamma_k \geq 0 \qquad \text{and} \qquad \gamma_k \to 0 \quad \text{as } k \to \infty.$$

- Let ξ be a **random variable** with values in

$$X := \{x \in \ell^\infty : \lim_{k\to\infty} x_k = 0\} \subset \ell^\infty$$

  whose **components** ξ_k are independent and each **uniformly distributed** on the interval $[-\gamma_k, \gamma_k]$.

Definition

Let $\mathcal{U}_E$ denote the **probability distribution** of ξ.

Strong and generalised modes of $\mathcal{U}_E$

Theorem

If there is an $m \in \mathbb{N}$ with

$$|x_m| = \gamma_m > 0,$$

then $x \in X$ is **not a strong mode** of $\mathcal{U}_E$.

Theorem

A point $x \in X$ is a **generalised mode** of $\mathcal{U}_E$ if and only if

$$|x_k| \leq \gamma_k \quad \text{for all } k \in \mathbb{N}.$$

Set-up

Assumptions

1. **Prior distribution** $\mu_0 := \mathcal{U}_E$ on
$$X = \left\{ x \in \ell^\infty : \lim_{k\to\infty} x_k = 0 \right\}.$$
2. For given **data** y, **posterior distribution** μ^y on X is given by
$$\mu^y(\mathrm{d}x) = \frac{1}{Z} \exp(-\Phi(x))\mu_0(\mathrm{d}x).$$

Objective functional

Conjecture

Generalised modes of μ^y can be characterised as **minimisers** of $\Phi\colon X \to \mathbb{R}$ in

$$E := \{x \in X : |x_k| \leq \gamma_k \text{ for all } k \in \mathbb{N}\}.$$

This is equivalent to **minimising** $I\colon X \to \mathbb{R} \cup \{\infty\}$,

$$I(x) := \Phi(x) + \iota_E(x),$$

where ι_E denotes the **indicator function**

$$\iota_E(x) := \begin{cases} 0 & \text{if } x \in E, \\ \infty & \text{otherwise.} \end{cases}$$

Variational characterisation of GMAP estimates

Assumption

The function $\Phi\colon X \to \mathbb{R}$ is **Lipschitz continuous on bounded sets**, i.e., for every $r > 0$, there exists $L_r > 0$ such that

$$|\Phi(x_1) - \Phi(x_2)| \le L_r \|x_1 - x_2\|_X \quad \text{for all } x_1, x_2 \in B^r(0).$$

Theorem

A point $\hat{x} \in X$ is a **generalised mode** of μ^y if and only if it is a **minimiser** of $I\colon X \to \mathbb{R} \cup \{\infty\}$,

$$I(x) := \Phi(x) + \iota_E(x).$$

Here, **generalised MAP estimator** corresponds to **Ivanov regularisation** with compact set E.

Set-up for consistency of generalised MAP estimator

Bayesian inverse problems

Governed by ill-posed operator equation

$$y = F(x) + \varepsilon\eta.$$

1. **Finite-dimensional data** $y \in \mathbb{R}^d$.
2. **Additive Gaussian noise**, $\eta \sim \mathcal{N}(0, \Sigma)$, noise level $\varepsilon > 0$.
3. Prior distribution $\mu_0 := \mathcal{U}_E$ on $X := \{x \in \ell^\infty : \lim_{k\to\infty} x_k = 0\}$.
4. Posterior distribution

$$\mu^y(\mathrm{d}x) = \frac{1}{Z}\exp\Big(-\underbrace{\frac{1}{2\varepsilon^2}\Big\|\Sigma^{-\frac{1}{2}}(F(x)-y)\Big\|^2}_{=\Phi(x)}\Big)\mu_0(\mathrm{d}x).$$

Set-up

Goal: Show **consistency** of generalised MAP estimator in **small noise limit** and **frequentist set-up**.

Assumptions

1. **True solution** $x^\dagger \in X$ exists.
2. **Sequence** $\{y_n\}_{n\in\mathbb{N}}$ **of measurements**, given by

$$y_n = F(x^\dagger) + \varepsilon_n \eta_n.$$

3. $\varepsilon_n \to 0$ as $n \to \infty$.
4. $\eta_n \sim \mathcal{N}(0, \Sigma)$ are i.i.d. Gaussian random variables.

For every $n \in \mathbb{N}$, let $x_n \in X$ be a **minimiser** of

$$I_n(x) := \frac{1}{2\varepsilon_n^2} \left\| \Sigma^{-\frac{1}{2}}(F(x) - y_n) \right\|^2 + \iota_E(x).$$

Consistency

Theorem

Suppose that $x^\dagger \in E$ and that F is **closed**. Then $\{x_n\}_{n\in\mathbb{N}}$ contains a **convergent subsequence** whose limit $\bar{x} \in X$ satisfies

$$F(\bar{x}) = F(x^\dagger) \quad \text{almost surely.}$$

Corollary

If, in addition, F is **injective**, then

$$x_n \to x^\dagger \quad \text{in probability as } n \to \infty.$$

Conclusion

- We have established **conditions** for the **coincidence** of strong modes and generalised modes.
- For priors with strictly bounded and uniformly distributed components, **generalised MAP estimates** are given as **minimisers** of a canonical objective functional.
- The **generalised MAP estimator** based upon such priors is **consistent** for nonlinear inverse problems with additive Gaussian noise.

References

C. Clason, T. Helin, R. Kretschmann, and P. Piiroinen.
Generalized modes in Bayesian inverse problems.
SIAM/ASA Journal on Uncertainty Quantification, 7(2):625–684, 2019.

M. Dashti, K. J. Law, A. M. Stuart, and J. Voss.
MAP estimators and their consistency in Bayesian nonparametric inverse problems.
Inverse Problems, 29(9):095017, 2013.

R. Kretschmann.
Nonparametric Bayesian inverse problems with Laplacian noise.
Doctoral thesis, Universität Duisburg-Essen, 2019.

S. Agapiou, M. Burger, M. Dashti, and T. Helin.
Sparsity-promoting and edge-preserving maximum a posteriori estimators in non-parametric Bayesian inverse problems.
Inverse Problems, 34(4):045002, 2018.