GENERALISED MODES IN BAYESIAN INVERSE PROBLEMS

Christian Clason[1], Tapio Helin[2], **Remo Kretschmann**[1], and 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

Framework

Nonparametric Bayesian inference

1. **Unknown quantity** x in separable Banach space X with **prior distribution** μ_0.
2. **Noisy data** y in separable Hilbert space Y. For every $y \in Y$, the **posterior distribution** μ^y has a density w.r.t. the prior distribution
$$\frac{d\mu^y}{d\mu_0}(x) = \frac{\exp(-\Phi(x;y))}{Z(y)}$$
where $\Phi\colon X \times Y \to \mathbb{R}$ is the **negative log-likelihood** and
$$Z(y) := \int_X \exp(-\Phi(x;y))\,\mu_0(dx) \in (0,\infty).$$

In particular, **Bayesian inverse problems** described by ill-posed operator equation.

Goals of this work [2]

1. **Extend definition** of **modes** and corresponding **MAP estimates** in **nonparametric Bayesian inference** to cover cases where previous approach fails, such as priors that are not quasi-invariant along any direction.
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** based upon **priors with strictly bounded components** are given as **minimisers** of canonical objective functional.
4. Study **consistency** for **Bayesian inverse problems**.

Modes in Infinite-dimensional Spaces

In **finite-dimensional** spaces, **modes** of probability measure typically defined as **maximisers** of its **density** w.r.t. Lebesgue measure.

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

For this reason, **modes** in **infinite-dimensional** spaces typically defined via **asymptotic small ball probabilities** [3, 5].

Definition

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

Variational Characterisation of MAP Estimates

Let μ be a Borel probability measure on a separable Banach space X. An element $h \in X$ is called **admissible shift** if the shifted measure $\mu_h := \mu(\cdot - h)$ is equivalent to μ. Let H denote the **set of admissible shifts** for μ.

Definition

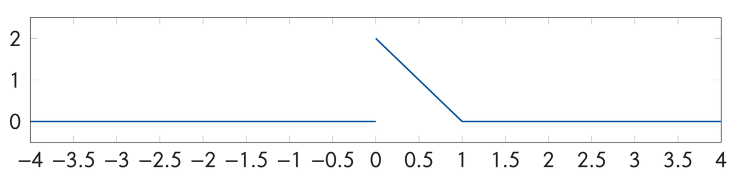
A function $I\colon H \to \mathbb{R}$ is called **Onsager–Machlup functional** for μ, if for all $h_1, h_2 \in H$ we have
$$\lim_{\delta\to 0}\frac{\mu(B^\delta(h_1))}{\mu(B^\delta(h_2))} = \exp(I(h_2) - I(h_1)).$$

Maximum a posteriori (MAP) estimates defined as **modes** of **posterior distribution** μ^y. Under certain conditions, MAP estimates are precisely **minimisers** of **Onsager–Machlup functional** I for μ^y in case of **Gaussian prior** [3, 4] or **Besov prior** [1] and
$$I(x) = \Phi(x;y) + R(x).$$

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

There are applications where **strict bounds** on the admissible values of the parameter x emerge in a natural way, e.g., radiography, electrical impedance tomography.

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

Definition

Let μ be a Borel probability measure on a separable Banach space X. 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 an **approximating sequence** $\{w_n\}_{n\in\mathbb{N}} \subset X$ such that $w_n \to \hat{x}$ in X and
$$\lim_{n\to\infty}\frac{\mu(B^{\delta_n}(w_n))}{\sup_{x\in X}\mu(B^{\delta_n}(x))} = 1.$$

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

Consider general Borel probability measure μ on separable Banach space X.

Theorem

Let $\hat{x} \in X$ be a **generalised mode** of μ. Then $\hat{x}$ is a **strong mode** if and only if for every sequence $\{\delta_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ with $\delta_n \to 0$, there exists an approximating sequence $\{w_n\}_{n\in\mathbb{N}} \subset X$ with $w_n \to \hat{x}$ and
$$\lim_{n\to\infty}\frac{\mu(B^{\delta_n}(\hat{x}))}{\mu(B^{\delta_n}(w_n))} = 1.$$

Corollary

Let $\hat{x} \in X$ be a generalised mode of μ. If there exists an $r > 0$ such that
$$\lim_{\delta\to 0}\frac{\mu(B^\delta(\hat{x}))}{\sup_{w\in B^r(\hat{x})}\mu(B^\delta(w))} = 1,$$
then $\hat{x}$ is a strong mode.

Idea: Characterise convergence of approximating sequence by **convergence rate**.

Theorem

Let $\hat{x} \in X$ be a **generalised mode** of μ. If

1. for every positive sequence $\{\delta_n\}_{n\in\mathbb{N}}$ with $\delta_n \to 0$, there exists an approximating sequence $\{w_n\}_{n\in\mathbb{N}}$ such that
$$\lim_{n\to\infty}\frac{\|w_n - \hat{x}\|_X}{\delta_n} = 0,$$
2. the family of functions $\{f_n\}_{n\in\mathbb{N}}$ on $[0,1]$ defined by
$$f_n : [0,1] \to \mathbb{R}, \qquad f_n(r) := \frac{\mu(B^{r(\delta_n + \|w_n - \hat{x}\|_X)}(\hat{x}))}{\mu(B^{\delta_n + \|w_n - \hat{x}\|_X}(\hat{x}))},$$
is equicontinuous at $r = 1$,

then $\hat{x}$ is a **strong mode**.

Idea: Characterise convergence of approximating sequence in **subspace topology**.

Theorem

Suppose that the **space of admissible shifts** H possesses a **dense continuously embedded subspace** $(E, \|\cdot\|_E) \subset H$ such that for every $h \in E$ the **density** of μ_h w.r.t. μ has a **continuous representative** $\frac{d\mu_h}{d\mu} \in C(X)$. Let $\hat{x} \in X$ be a **generalised mode** of μ. If

1. for every $\{\delta_n\} \subset (0,\infty)$ with $\delta_n \to 0$ there is an approximating sequence $\{w_n\}_{n\in\mathbb{N}} \subset \hat{x} + E$ with $\|w_n - \hat{x}\|_E \to 0$,
2. there is an $R > 0$ such that
$$f_R : (E, \|\cdot\|_E) \to \mathbb{R}, \qquad f_R(h) := \sup_{x\in B^R(\hat{x})}\left|\frac{d\mu_h}{d\mu}(x) - 1\right|$$
is continuous at $h = 0$,

then $\hat{x}$ is a **strong mode**.

Corollary

Let $\hat{x} \in X$ be a generalised mode of μ that satisfies condition 1. If additionally
$$\lim_{h\to_E 0}\frac{d\mu_h}{d\mu}(\hat{x}) = 1$$
and there is an $r > 0$ such that the family
$$\left\{\frac{d\mu_h}{d\mu} : h \in E, \|h\|_E < r\right\}$$
is equicontinuous in $\hat{x}$, then $\hat{x}$ is a strong mode.

Modes of Uniform Prior Distribution

Idea: Define probability measure on separable subspace of ℓ^∞ whose mass is concentrated on set of sequences with **strictly bounded components**.

Set
$$X := \overline{\operatorname{span}\{e_n\}_{n\in\mathbb{N}}} = \left\{x \in \ell^\infty : \lim_{k\to\infty} x_k = 0\right\} \subset \ell^\infty,$$
where $\{e_n\}_{n\in\mathbb{N}}$ denotes the **standard unit vectors** in ℓ^∞, i.e., $\{e_n\}_k = 1$ for $n = k$ and 0 otherwise. Then, X equipped with $\|x\|_\infty := \sup_{k\in\mathbb{N}}|x_k|$ is a separable Banach space.

Definition

For a **given sequence** $\{\gamma_n\}_{n\in\mathbb{N}}$ with $\gamma_k \geq 0$ for all $k \in \mathbb{N}$ and $\gamma_k \to 0$ define the X-valued random variable
$$\xi := \sum_{k=1}^\infty \gamma_k \xi_k e_k,$$
where $\{\xi_k\}_{k\in\mathbb{N}}$ are i.i.d. real-valued random variables, each **uniformly distributed** on $[-1,1]$. Then, define the **probability measure** μ_γ on X by
$$\mu_\gamma(A) := \mathbb{P}[\xi \in A] \quad \text{for all } A \in \mathcal{B}(X).$$

Question: What are the strong and generalised modes of μ_γ?

Define
$$E_\gamma := \{x \in X : |x_k| \leq \gamma_k \text{ for all } k \in \mathbb{N}\},$$
$$E_\gamma^0 := \{x \in X : |x_k| < \gamma_k \text{ for all } k \in \mathbb{N}, x_k \neq 0 \text{ for finitely many } k \in \mathbb{N}\}.$$

Theorem

1. Every point $x \in E_\gamma^0$ **is a strong mode** of μ_γ.
2. If there is an $m \in \mathbb{N}$ with $|x_m| = \gamma_m > 0$, then x is **not a strong mode** of μ_γ.

Proposition

1. There are $\gamma \in X$ and $x \in E_\gamma \setminus E_\gamma^0$ with $|x_k| < \gamma_k$ for all $k \in \mathbb{N}$ such that x **is a strong mode** of μ_γ.
2. There are $\gamma \in X$ and $x \in E_\gamma \setminus E_\gamma^0$ with $|x_k| < \gamma_k$ for all $k \in \mathbb{N}$ such that x is **not a strong mode** of μ_γ.

Theorem

A point $x \in X$ is a **generalised mode** of μ_γ if and only if $x \in E_\gamma$.

Variational Characterisation of Generalised MAP Estimates

Bayesian inference

1. **Uniform prior distribution** $\mu_0 := \mu_\gamma$ on $X := \{x \in \ell^\infty : \lim_{k\to\infty} x_k = 0\}$.
2. Fix $y \in Y$. **Posterior distribution** given by
$$\mu^y(dx) = \frac{1}{Z}\exp(-\Phi(x))\,\mu_0(dx).$$
3. The function $\Phi\colon X \to \mathbb{R}$ is **Lipschitz continuous on bounded sets**, i.e., for every $r > 0$, there exists $L = L_r > 0$ such that for all $x_1, x_2 \in B^r(0)$ we have
$$|\Phi(x_1) - \Phi(x_2)| \leq L\|x_1 - x_2\|_X.$$

Generalised MAP estimates defined as **generalised modes** of posterior distribution μ^y.

Goal: Characterise generalised MAP estimates as **minimisers** of appropriate objective functional.

Onsager–Machlup functional not defined for **prior distribution** μ_0, but **generalised modes** of μ_0 are precisely **minimisers** of **indicator function** $\iota_{E_\gamma}\colon X \to \overline{\mathbb{R}} := \mathbb{R} \cup \infty$,
$$\iota_{E_\gamma} := \begin{cases} 0 & \text{if } x \in E_\gamma,\\ \infty & \text{otherwise.}\end{cases}$$

Define $I\colon X \to \overline{\mathbb{R}}$,
$$I(x) := \Phi(x) + \iota_{E_\gamma}(x).$$

Proposition (generalised Onsager–Machlup property)

Let $x_1, x_2 \in E_\gamma$ and $\{w_1^\delta\}_{\delta>0}, \{w_2^\delta\}_{\delta>0} \subset E_\gamma$ such that

1. $w_1^\delta \to x_1$ and $w_2^\delta \to x_2$ as $\delta \to 0$,
2. $w_1^\delta, w_2^\delta \in E_\gamma^\delta$ for all $\delta > 0$, where
$$E_\gamma^\delta := \{x \in X : |x_k| \leq \max\{\gamma_k - \delta, 0\} \text{ for all } k \in \mathbb{N}\} \subset E_\gamma.$$

Then,
$$\lim_{\delta\to 0}\frac{\mu^y(B^\delta(w_1^\delta))}{\mu^y(B^\delta(w_2^\delta))} = \exp(I(x_2) - I(x_1)).$$

Main theorem

Suppose that Φ is **Lipschitz continuous on bounded sets**. Then, a point $\hat{x} \in X$ is a **generalised MAP estimate** for μ^y if and only if it is a **minimiser** of I.

Sketch of proof

1. For every $\delta > 0$, let x^δ denote a **maximiser** of
$$x \mapsto \mu^y(B^\delta(x)).$$
For every positive sequence $\{\delta_n\}_{n\in\mathbb{N}}$ with $\delta_n \to 0$, the sequence $\{x^{\delta_n}\}_{n\in\mathbb{N}}$ contains a **subsequence** that **converges strongly** in X to some $\bar{x} \in E_\gamma$.
2. Any **cluster point** $\bar{x} \in E_\gamma$ of $\{x^{\delta_n}\}_{n\in\mathbb{N}}$ is a **minimiser** of I.
3. Use 1, 2, Lipschitz continuity and generalised OM property to show proposition.

For inverse problems subject to Gaussian noise, **generalised MAP estimator** coincides with **Ivanov regularisation** using compact set E_γ. **Minimisers** generically lie **on boundary** of compact set if Ivanov functional is convex [6].

Consistency for Inverse Problems with Gaussian Noise

Bayesian inverse problems with **uniform prior distribution** $\mu_0 = \mu_\gamma$, **finite-dimensional data** $y \in Y := \mathbb{R}^K$, and **additive Gaussian noise**, governed by operator equation
$$y = F(x) + \delta\eta.$$

1. **Nonlinear operator** $F\colon X \to Y$.
2. **Gaussian noise** $\eta \sim \mathcal{N}(0,\Sigma)$ with positive definite covarance matrix $\Sigma \in \mathbb{R}^{K\times K}$, **scaled** by $\delta > 0$. **Negative log-likelihood** given by
$$\Phi(x;y) = \frac{1}{2\delta^2}\|\Sigma^{-\frac{1}{2}}(F(x) - y)\|_Y^2.$$

Goal: Show consistency in **small noise limit** in **frequentist setup**.

True solution $x^\dagger \in E_\gamma$, **sequence** $\{y_n\}_{n\in\mathbb{N}} \subset Y$ of **measurements** given by
$$y_n = F(x^\dagger) + \delta_n\eta_n,$$
where $\delta_n \to 0$ and $\eta_n \sim \mathcal{N}(0,\Sigma)$ are i.i.d. **Gaussian random variables**.

Theorem

Suppose that $F\colon X \to Y$ is **closed** and $x^\dagger \in E_\gamma$. For every $n \in \mathbb{N}$, let $x_n \in X$ be a **minimiser** of
$$I^{y_n}(x) := \frac{1}{2\delta^2}\|\Sigma^{-\frac{1}{2}}(F(x) - y_n)\|_Y^2 + \iota_{E_\gamma}(x).$$
Then, $\{x_n\}_{n\in\mathbb{N}}$ contains a **convergent subsequence** whose limit $\bar{x} \in E_\gamma$ satisfies $F(\bar{x}) = F(x^\dagger)$ almost surely.

Corollary

If F is **injective**, then $x_n \to x^\dagger$ in probability as $n \to \infty$.

References

[1] 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.

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

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

[4] R. Kretschmann. *Nonparametric Bayesian Inverse Problems with Laplacian Noise*. PhD thesis, Universität Duisburg-Essen, 2019. Submitted.

[5] H. C. Lie and T. J. Sullivan. Equivalence of weak and strong modes of measures on topological vector spaces. *Inverse Problems*, 34(11):115013, 2018.

[6] A. Neubauer and R. Ramlau. On convergence rates for quasi-solutions of ill-posed problems. *Electron. Trans. Numer. Anal.*, 41:81–92, 2014.