Strong maximum a posteriori estimation in Bayesian inverse problems

Hefin Lambley
Mathematics Institute, University of Warwick

5th December 2023

velocity field

radiodensity field

unblurred image

$\mathcal{G}\colon X \to Y$

observation operator

Lagrangian data

CT measurements

pixels for blurred image

unknown parameter x
in separable Banach space X

observation $\mathcal{G}(x)$
in separable Banach space Y

Inverse problem

Given a (possibly corrupted) observation $y \in Y$ of $\mathcal{G}(x)$, recover $x \in X$.

Inverse problems are usually ill-posed, so minimising the misfit

$$\operatorname*{arg\,min}_{x \in X} \Phi(x; y), \qquad \Phi(x; y) := \frac{1}{2}\|\mathcal{G}(x) - y\|_Y^2$$

is unstable and affected by observational noise.

Classical approaches to inverse problems use regularisation

[Tikhonov, 1943; Engl, Hanke & Neubauer 1996]

Tikhonov regularisation — Pick a penalty functional $\Omega \colon X \to [0, +\infty]$, a regularisation parameter $\alpha > 0$, and minimise

$$I(x; y) = \Phi(x; y) + \alpha\Omega(x).$$

Theorem

Under suitable functional-analytic assumptions on X, Y, $\mathcal{G}$, and Ω, there exists a minimiser of I which stably depends on the data y.

Adding "prior information" through Ω restores well-posedness!

Prior information can be naturally integrated in the Bayesian approach

[Stuart, 2010]

Additive-noise Bayesian inverse problem

Infer $x \in X$ from data $y \in Y$ given the model

$$y = \mathcal{G}(x) + \xi,$$

$$\xi \sim \tau_0,$$

$$x \sim \mu_0.$$

Model $y = \mathcal{G}(x) + \xi,$
$\xi \sim \tau_0,$
$x \sim \mu_0.$

Theorem (Bayes' rule for functions) [Dashti & Stuart, 2017]

Assume:

- mild measurability and integrability assumptions;
- $\tau_{\mathcal{G}(x)} = \tau_0(\cdot - \mathcal{G}(x))$ is absolutely continuous wrt. τ_0 for μ_0-almost all x.

Then the distribution μ^y of $x \mid y$ is absolutely continuous with respect to μ_0, and

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(x) = \frac{1}{Z(y)} \exp\bigl(-\Phi(x;y)\bigr), \quad \Phi(x;y) = -\log \frac{\mathrm{d}\tau_{\mathcal{G}(x)}}{\mathrm{d}\tau_0}(y).$$

Example: observations corrupted by additive Gaussian noise

Model $y = \mathcal{G}(x) + \xi, \quad \xi \sim N(0, \Sigma).$

Prior Centred nondegenerate Gaussian measure μ_0.

Posterior Given by Bayes' rule:

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(x) = \frac{1}{Z(y)} \exp\Big(-\Phi(x;y)\Big), \quad \Phi(x;y) = \frac{1}{2}\left\|\Sigma^{-1/2}\Big(\mathcal{G}(x) - y\Big)\right\|^2.$$

Tikhonov regularisation

Regularisation functional Ω and misfit term

Minimiser $x^\dagger$ of Tikhonov functional

Bayesian inference

Prior μ_0 and likelihood $y \mid x$

Posterior μ^y given by Bayes' rule

Maximum a posteriori (MAP) estimator for absolutely continuous μ^y on $X = \mathbb{R}^d$

A MAP estimator is a maximiser of the posterior density $\rho^y : \mathbb{R}^d \to [0, +\infty]$.

When $X = \mathbb{R}^d$, $Y = \mathbb{R}^n$, $\mu_0 \sim N(0, \Theta)$, $\xi \sim N(0, \Sigma)$,

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mathcal{L}^d}(x; y) \propto \exp\left(-\frac{1}{2}\left\|\Sigma^{-1/2}(\mathcal{G}(x) - y)\right\|^2 - \frac{1}{2}\left\|\Theta^{-1/2}x\right\|^2\right),$$

so a maximiser of the density minimises the Tikhonov functional

$$I(x; y) = \frac{1}{2}\left\|\Sigma^{-1/2}(\mathcal{G}(x) - y)\right\|^2 + \frac{1}{2}\left\|\Theta^{-1/2}x\right\|^2.$$

Question: how can we extend this connection beyond the finite-dimensional setting?

Modes can be defined using small-ball probabilities

[Dashti, Law, Stuart & Voss, 2013]

Assumptions X is a separable Banach space and μ is a Borel probability measure on X.

Idea Look for points which asymptotically attain the supremal ball mass

$$M_r = \sup_{x \in X} \mu(B_r(x)).$$

Definition: strong mode

Any point $x^\star \in \operatorname{supp}(\mu)$ such that

$$\lim_{r \to 0} \frac{M_r}{\mu(B_r(x^\star))} = 1.$$

Modes can be defined using small-ball probabilities

[Helin & Burger, 2015; Ayanbayev, Klebanov, Lie & Sullivan, 2022a]

Another approach is to compare balls around $x^\star$ with those around any other $x \in X$.

Definition: (global) weak mode

Any point $x^\star \in \operatorname{supp}(\mu)$ such that, for all $x \in X$,

$$\limsup_{r \to 0} \frac{\mu(B_r(x))}{\mu(B_r(x^\star))} \leq 1.$$

This also admits an order-theoretic interpretation. [L. & Sullivan, 2023]

Proposition

Any strong mode is a weak mode.

Proposition [Lie & Sullivan, 2018; Ayanbayev, Klebanov, Lie, Sullivan, 2022a]

There exist measures with a weak mode that is not strong.

"Strong" or "weak" is really a regularity condition on the measure.

Proposition: strong–weak dichotomy [L., 2023]

If μ has a strong mode, then all weak modes are strong modes.

Definition: Onsager–Machlup functional

Let $\emptyset \neq E \subseteq X$. Then $I \colon E \to \mathbb{R}$ is an OM functional for μ if

$$\lim_{r \to 0} \frac{\mu(B_r(x))}{\mu(B_r(y))} = \exp\Big(I(y) - I(x)\Big) \quad \text{for all } x, y \in E.$$

Definition: property $M(\mu, E)$ [Ayanbayev, Klebanov, Lie & Sullivan, 2022a]

Property $M(\mu, E)$ holds if there exists $x^\star \in E$ such that, for all $x \notin E$,

$$\lim_{r \to 0} \frac{\mu(B_r(x))}{\mu(B_r(x^\star))} = 0.$$

Theorem [Ayanbayev, Klebanov, Lie & Sullivan 2022a]

Assume μ admits an OM functional $I\colon E \to \mathbb{R}$ and property $M(\mu, E)$ holds. Then:

$$x^\star \text{ is a weak mode} \iff x^\star \text{ minimises } I.$$

Theorem

Assume μ_0 has OM functional $I_0\colon E \to \mathbb{R}$ and property $M(\mu_0, E)$ holds.
Suppose $\mu^y \ll \mu_0$ and, for some continuous potential Φ, has density

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(x) = \exp\Big(-\Phi(x)\Big).$$

Then μ^y has OM functional $I^y(x) = I_0(x) + \Phi(x)$ and property $M(\mu^y, E)$ holds.

Example: Gaussian measures on separable Hilbert spaces

Let γ be a centred Gaussian on a separable Hilbert space X. The measure is characterised by its covariance operator $\mathcal{C} \colon X \to X$, which can be written as

$$\mathcal{C}x = \sum_{n \in \mathbb{N}} \sigma_n^2 \langle x, e_n \rangle e_n, \quad (e_n)_{n \in \mathbb{N}} \text{ orthonormal basis of } X, \quad (\sigma_n^2)_{n \in \mathbb{N}} \text{ decreasing.}$$

We can treat γ as a product measure in the basis $(e_n)_{n \in \mathbb{N}}$:

$$\gamma = \bigotimes_{n \in \mathbb{N}} N(0, \sigma_n^2).$$

The *Cameron–Martin space* E consists of all $h \in X$ for which

$$\|h\|_E^2 \coloneqq \sum_{n \in \mathbb{N}} \sigma_n^{-2} \langle h, e_n \rangle_X^2 < \infty.$$

It is classical that γ has OM functional $I(x) = \frac{1}{2} \|x\|_E^2$ defined on E,
and property $M(\gamma, E)$ is implicit in results of Dashti, Law, Stuart & Voss (2013).

With the definition established, how can we prove Bayesian inverse problems have MAP estimators?

We restrict to the following class of measures, motivated by Bayesian posteriors.

Problem statement

Let X be a separable Banach space and let μ_0 be a centred nondegenerate Gaussian. Suppose that, for some continuous potential $\Phi\colon X \to \mathbb{R}$,

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(x) = \exp\Big(-\Phi(x)\Big).$$

When does μ^y have a strong MAP estimator?

We just need to prove the existence of at least one strong mode. Then:

- strong and weak modes coincide;
- if the M-property also holds, then strong modes coincide with minimisers of I^y.

Theorem [Dashti, Law, Stuart & Voss, 2013]

Suppose also that X is Hilbert and that Φ is bounded below and locally Lipschitz. Then μ^y has a strong mode.

Theorem [Kretschmann 2019, 2023]

Suppose X is Hilbert and that Φ is locally Lipschitz and satisfies the *lower cone condition*, i.e. there exists $L > 0$ such that

$$\Phi(x) \geq \Phi(0) - L\|x\|_X \text{ for all } x \in X.$$

Then μ^y has a strong mode.

Theorem [Klebanov & Wacker, 2023]

Let $X = \ell^p(\mathbb{N};\mathbb{R})$, $1 \leq p < \infty$, and let μ_0 be a diagonal Gaussian measure:

$$\mu_0 = \bigotimes_{n\in\mathbb{N}} N(0, \sigma_n^2).$$

Suppose that Φ is bounded below and locally Lipschitz. Then μ^y has a strong mode.

Proof strategy

1. Take an *asymptotic maximising family* (AMF) $(x_r)_{r>0}$ for μ^y, i.e. a net satisfying

$$\mu^y(B_r(x_r)) > \big(1 - \varepsilon(r)\big) \sup_{x \in X} \mu^y(B_r(x)) \quad \text{for some } \varepsilon(r) \to 0.$$

2. Show that $(x_r)_{r>0}$ has a bounded subsequence using an *explicit Anderson inequality*.

Theorem: explicit Anderson inequality [Dashti, Law, Stuart & Voss, 2013]

Let X be Hilbert and let γ be a centred Gaussian. There exists $a > 0$ such that

$$\frac{\gamma(B_r(x))}{\gamma(B_r(0))} \leq \exp\Big(a\big(r^2 - (\|x\| - r)^2\big)\Big) \qquad \text{for any } x \in X \text{ and } r > 0.$$

Using density for μ^y and that $(x_r)_{r>0}$ is an AMF for μ^y, can show

$$0 < \liminf_{r \to 0} \frac{\mu_0(B_r(x_r))}{\mu_0(B_r(0))} \leq \liminf_{r \to 0} \exp\Big(a\big(r^2 - (\|x_r\| - r)^2\big)\Big).$$

3. Extract weakly convergent subsequence of $(x_r)_{r>0}$ and show convergence is also strong.
4. Use the regularity of μ^y to show the limit point is a strong mode.

Theorem [L., 2023]

Let X be any separable Banach space, and suppose Φ is continuous and for each $\eta > 0$, there exists $K(\eta) \in \mathbb{R}$ such that

$$\Phi(x) \geq K(\eta) - \eta \|x\|_X^2.$$

Then μ^y has a strong mode.

To prove this, we upgrade a couple of steps in the proof strategy.

1. Take an AMF $(x_r)_{r>0}$ for μ^y.
2. Approximate it by an AMF lying in E with a bounded subsequence ~~in X~~ in E.
3. Extract ~~X-weakly~~ E-weakly convergent subsequence, and show convergence is X-strong.
4. Use regularity of μ^y to show the limit point is a strong mode.

The original explicit Anderson inequality is of the form

$$\frac{\gamma(B_r(x))}{\gamma(B_r(0))} \leq \exp\Big(a\big(r^2 - (\|x\| - r)^2\big)\Big) \quad \text{for any } x \in X \text{ and } r > 0.$$

This complicates matters when X doesn't have Hilbert structure.

Theorem: explicit Anderson inequality in Cameron–Martin norm

Suppose X is a separable Banach space and γ is a centred nondegenerate Gaussian. Then

$$\frac{\gamma(B_r(x))}{\gamma(B_r(0))} \leq \exp\Big(-\frac{1}{2} \min_{h \in B_r(x) \cap E} \|h\|_E^2\Big) \quad \text{for any } x \in X \text{ and } r > 0.$$

A similar result can be found in Ghosal & van der Vaart (2017).

Corollary: this also proves property $M(\gamma, E)$ in separable Banach spaces.

Suppose X is a separable Banach space and γ is a centred nondegenerate Gaussian. Then

$$\frac{\gamma(B_r(x))}{\gamma(B_r(0))} \leq \exp\Big(-\frac{1}{2} \min_{h \in B_r(x) \cap E} \|h\|_E^2\Big) \quad \text{for any } x \in X \text{ and } r > 0.$$

Using the inequality, we can show any AMF is approximated by another with a bounded subsequence in the Hilbert space E.

We can therefore extract an E-weakly convergent subsequence and push it through the compact embedding of E in X.

This extends the correspondence between MAP estimators and minimisers of a Tikhonov functional to Banach spaces

Corollary

When X is a separable Banach space and μ_0 is a centred nondegenerate Gaussian, the posterior μ^y of the Bayesian inverse problem

$$y = \mathcal{G}(x) + \xi, \qquad \xi \sim N(0, \Sigma), \qquad x \sim \mu_0$$

has at least one strong MAP estimator, and strong MAP estimators are minimisers of

$$I^y(x) = \frac{1}{2} \left\| \Sigma^{-1/2} (\mathcal{G}(x) - y) \right\|^2 + \frac{1}{2} \|x\|_E^2.$$

The analysis for non-Gaussian priors seems much harder — no general theory.
Similar results do exist for Besov priors. [Agapiou, Burger, Dashti & Helin, 2018]

In inverse problems, MAP estimators connect classical and Bayesian approaches.
This is well known in finite dimensions, but hard to prove in the nonparametric setting.

Modes of probability measures can be defined in very general settings using small balls.
This overcomes a lack of posterior Lebesgue density in infinite dimensions.

Bayesian inverse problems with Gaussian priors on Banach spaces have MAP estimators.
The proof exploits an explicit Anderson inequality in Cameron–Martin norm
and there is much work to be done beyond the Gaussian case.

Thank you!

Slides and paper available at
`warwick.ac.uk/htlambley`

References

Ayanbayev, Klebanov, Lie & Sullivan (2022a). Γ-convergence of Onsager–Machlup functionals: I. With applications to maximum a posteriori estimation in Bayesian inverse problems. Inverse Probl. 38.2, 025005, doi:10.1088/1361-6420/ac3f81.

Agapiou, Burger, Dashti & Helin (2018). Sparsity-promoting and edge-preserving maximum a posteriori estimators in non-parametric Bayesian inverse problems. Inverse Probl. 34.4, 045002, doi:10.1088/1361-6420/aaacac.

Dashti, Law, Stuart & Voss (2013). MAP estimators and their consistency in Bayesian nonparametric inverse problems. Inverse Probl. 29.9, 095017, doi:10.1088/0266-5611/29/9/095017.

Dashti & Stuart (2017). The Bayesian approach to inverse problems. In *Handbook of Uncertainty Quantification*, Springer, pp. 311–428, doi:10.1007/978-3-319-12385-1_7.

Engl, Hanke & Neubauer (1996). Regularization of inverse problems. Vol. 375 of Mathematics and Its Applications, Springer, ISBN 9780792341574.

Ghosal & van der Vaart (2017). Fundamentals of nonparametric Bayesian inference. CUP, doi:10.1017/9781139029834.

Helin & Burger (2015). Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. Inverse Probl. 31.8, 085009, doi:10.1088/0266-5611/31/8/085009.

Klebanov & Wacker (2023). Maximum a posteriori estimators in ℓ^p are well-defined for diagonal Gaussian priors. Inverse Probl. 39.6, 065009, doi:10.1088/1361-6420/acce60.

Kretschmann (2019). Nonparametric Bayesian inverse problems with Laplacian noise. PhD thesis, Universität Duisberg-Essen, doi:10.17185/duepublico/70452.

Kretschmann (2023). Are minimizers of the Onsager–Machlup functional strong posterior modes? SIAM/ASA JUQ 11.4, pp. 1105–1138, doi:10.1137/23M1546579.

Lambley (2023). Strong maximum a posteriori estimation in Banach spaces with Gaussian priors. Inverse Probl. 39.12, 125010, doi:10.1088/1361-6420/ad07a4.

Lambley & Sullivan (2023). An order-theoretic perspective on modes and maximum a posteriori estimation in Bayesian inverse problems. SIAM/ASA JUQ 11.4, pp. 1195–1224, doi:10.1137/22M154243X.

Lie & Sullivan (2018). Equivalence of weak and strong modes of measures on topological vector spaces. Inverse Probl. 34.11, 115013, doi:10.1088/1361-6420/aadef2.

Stuart (2010). Inverse problems: a Bayesian perspective. Acta Numer. 19, pp. 451–559, doi:10.1017/S0962492910000061

Tikhonov (1943). On the stability of inverse problems [in Russian]. Dokl. Akad. Nauk. SSSR 39, pp. 195–198.