Γ-convergence of Onsager–Machlup functionals and MAP estimation in non-parametric Bayesian inverse problems

B. Ayanbayev[1] I. Klebanov[2] H. C. Lie[3] **T. J. Sullivan**[1,4]

Numerical Analysis and Scientific Computing Seminar
University of Manchester, UK, 6 May 2022

[1] **University of Warwick, UK**
[2] Freie Universität Berlin, DE
[3] Universität Potsdam, DE
[4] **Alan Turing Institute, UK**

Motivation

- In applications such as the Bayesian approach to inverse problems and the analysis of transitions of dynamical systems, it is often desirable to summarise a complicated probability measure μ on a high-dimensional space X by a single point $x^{\star} \in X$ — a "point of maximum probability under μ".

Motivation

- In applications such as the Bayesian approach to inverse problems and the analysis of transitions of dynamical systems, it is often desirable to summarise a complicated probability measure μ on a high-dimensional space X by a single point $x^\star \in X$ — a "point of maximum probability under μ".
- Essentially, we are looking for **modes** of μ. In the Bayesian context, when μ is the posterior measure, these are **maximum a posteriori (MAP) estimators**.

Motivation

- In applications such as the Bayesian approach to inverse problems and the analysis of transitions of dynamical systems, it is often desirable to summarise a complicated probability measure μ on a high-dimensional space X by a single point $x^\star \in X$ — a "point of maximum probability under μ".
- Essentially, we are looking for **modes** of μ. In the Bayesian context, when μ is the posterior measure, these are **maximum a posteriori (MAP) estimators**.

Challenge I — Definition(s)

What does "point of maximum probability under μ" even mean when μ is a measure on a metric space X, with no uniform reference measure etc.?

Motivation

- In applications such as the Bayesian approach to inverse problems and the analysis of transitions of dynamical systems, it is often desirable to summarise a complicated probability measure μ on a high-dimensional space X by a single point $x^\star \in X$ — a "point of maximum probability under μ".
- Essentially, we are looking for **modes** of μ. In the Bayesian context, when μ is the posterior measure, these are **maximum a posteriori (MAP) estimators**.

Challenge I — Definition(s)

What does "point of maximum probability under μ" even mean when μ is a measure on a metric space X, with no uniform reference measure etc.?

Challenge II — Stability

Are such points stable under perturbations of μ, or perturbations of problem data determining μ (changes of prior, likelihood, data, discretisation. . .)?

Some notation

- Throughout, X will be a separable metric space— occasionally something better, such as a separable Banach or Hilbert space.
- $B_r(x) := \{y \in X \mid d(x, y) \leqslant r\}$ denotes the closed ball of radius $r \geqslant 0$ centred on $x \in X$.
- $\mathcal{P}(X)$ denotes the set of all probability measures on the Borel σ-algebra of X.
- (Separability ensures that every $\mu \in \mathcal{P}(X)$ has a non-empty support, i.e. there is some $x \in X$ with $\mu(B_r(x)) > 0$ for every $r > 0$, and so $M_r := \sup_{x \in X} \mu(B_r(x)) > 0$.)
- Given a positive sequence $\gamma = (\gamma_k)_{k \in \mathbb{N}}$, we have the corresponding weighted ℓ^p norm and weighted ℓ^p space:

$$\begin{aligned} \|h\|_{\ell^p_\gamma} &:= \left\|(h_k/\gamma_k)_{k \in \mathbb{N}}\right\|_{\ell^p}, \\ \ell^p_\gamma &:= \left\{h \in \mathbb{R}^{\mathbb{N}} \,\middle|\, (h_k/\gamma_k)_{k \in \mathbb{N}} \in \ell^p\right\}. \end{aligned}$$

Well-posedness of (Bayesian) inverse problems

- In an **inverse problem** we recover a parameter $u \in X$ from observed data $y \in Y$. Such problems are usually ill posed: the recovered $u^y \in X$ depends sensitively on y, and this sensitivity is worse the "nicer" the forward map $u \mapsto y$ is (and this is why inverse problems need to be regularised).
- In a **Bayesian inverse problem (BIP)**, the recovery of u from y is expressed in the form of a posterior probability distribution $\mu^y \in \mathcal{P}(X)$.
- BIPs *are* well-posed (Stuart, 2010; ...; Sprungk, 2020). The posterior μ^y is a stable function of the problem setup — the prior distribution $\mu_0 \in \mathcal{P}(X)$, the observed data $y \in Y$, and the likelihood model $\ell \colon X \to \mathcal{P}(Y)$ — with respect to e.g. the Hellinger, Kullback–Leibler, or Wasserstein distances on $\mathcal{P}(X)$, e.g.

$$\mathrm{KL}(\mu^y \| \mu^{y+\delta y}) \lesssim \|\delta y\| \quad \text{(for fixed } \ell \text{ and } \mu_0\text{)}.$$

- Are the MAP estimators, the "most likely points under μ^y", also stable?

Similarity of modes $\not\Rightarrow$ similarity of measures

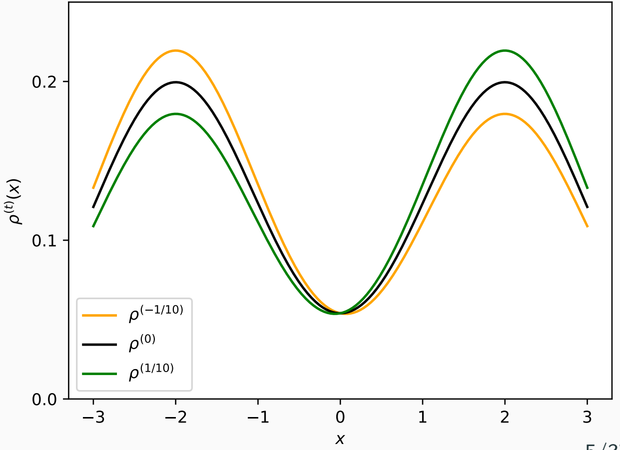
- Unfortunately, closeness of probability measures and closeness of their modes are "orthogonal" questions, even using a strong distance on $\mathcal{P}(X)$ like Kullback–Leibler.
- This is the case even for probability measures on $\mathbb{R}$ with continuous Lebesgue densities, for which a mode is easily defined as a maximiser of the density.
- Obviously, two measures can have very similar (or even the same) modes and yet be very different as measures, e.g. $\mathcal{N}(0, 1)$ and $\mathcal{N}(0, 10^6)$ or $\mu(E) = \int_E \max(0, 1 - |x|) \, \mathrm{d}x$!
- Perhaps if a sequence of probability measures converges "strongly enough", then their modes will also converge?
- Unfortunately, this is not the case.

Similarity of measures $\not\Rightarrow$ similarity of modes

Consider, for $t \in \mathbb{R}$, $\mu^{(t)} \in \mathcal{P}(\mathbb{R})$ with Lebesgue density

$$\rho^{(t)}(x) := \frac{(1+t)\exp(-\frac{1}{2}(x-r)^2) + (1-t)\exp(-\frac{1}{2}(x+r)^2)}{2\sqrt{2\pi}}.$$

- For $t > 0$ and for $r > 0$ large enough, $\rho^{(\pm t)}$ has a unique maximiser at $x^\star_{\pm t} \approx \pm r$.
- $\mathrm{KL}(\mu^{(t)} \| \mu^{(-t)}) \approx Ct^2$, and yet their modes are order 1 apart.
- This isn't too bad: The *cluster points* of the modes of $\mu^{(t)}$ as $t \to 0$ form the modes of $\mu^{(0)}$.

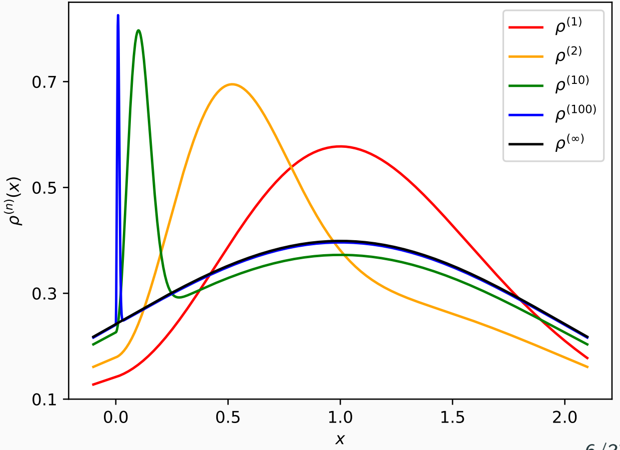


Similarity of measures $\not\Rightarrow$ similarity of modes

Consider, for $n \in \mathbb{N}$, $\mu^{(n)} \in \mathcal{P}(\mathbb{R})$ with Lebesgue density

$$\rho^{(n)}(x) := \frac{\exp\left(-\frac{1}{2}(x-1)^2\right) + \mathbb{1}[x \geqslant 0] 4n^2x^2 \exp(-n^2x^2)}{\sqrt{2\pi} + \sqrt{\pi}/n}$$

- Each $\rho^{(n)}$ has a unique maximiser at $x_n^\star \approx \frac{1}{n}$.
- Pointwise, $\rho^{(n)} \to \rho^{(\infty)}$, the density of $\mu^{(\infty)} = \mathcal{N}(1,1)$.
- $\mathrm{KL}(\mu^{(\infty)} \| \mu^{(n)}) \approx \frac{1}{n}$
- But the maximiser of $\rho^{(\infty)}$ is at $x_\infty^\star = 1 \neq \lim_{n\to\infty} x_n^\star$!



Ouch.

A role for Γ-convergence

- Evidently, these densities are not converging the "the right way", and even "strong" distances on $\mathcal{P}(X)$ like Kullback–Leibler are not the right notion of convergence.
- Modes are characterised as maximisers of the density — or minimisers of the negative log-density.
- The well-established notion of Γ-convergence from the calculus of variations (De Giorgi and Franzoni, 1975), which aims to give conditions for convergence of minimisers of minimisation problems, would seem to be a natural thing to try.
- (And it would be nice not to have to talk about densities, because not every measure has one...)

Overview

1. Modes, MAP estimators, and Onsager–Machlup functionals
2. Γ-convergence: A capsule summary
3. Γ-convergence of OM functionals
4. Bayesian inverse problems
5. MAP estimation for BIPs
6. Closing remarks

Modes, MAP estimators, and Onsager–Machlup functionals

Defining a mode of a measure. 1: Strong modes

- There is no such thing as a "Lebesgue-like" uniform reference measure λ on an infinite-dimensional space X (Sudakov, 1959), so we can't define a mode of μ as a maximiser of the density $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$.
- Over the last decade, it has become common to define modes directly using the masses of metric balls in the small-radius limit (Dashti et al., 2013; Helin and Burger, 2015; Clason et al., 2019).

Definition 1 (after Dashti et al. (2013))

A **strong mode** of $\mu \in \mathcal{P}(X)$ is any $x^\star \in X$ such that

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

where $B_r(x) := \{x' \in X \mid d(x, x') \leqslant r\}$ and $M_r := \sup_{x \in X} \mu(B_r(x))$.

Defining a mode of a measure. 2: Weak modes

Note that $\mu(B_r(x^\star)) \,/\, M_r \in [0, 1]$, so

$$
\begin{aligned}
x^\star \text{ is a strong mode} &\iff \lim_{r \to 0} \frac{\mu(B_r(x^\star))}{M_r} = 1 \\
&\iff \liminf_{r \to 0} \frac{\mu(B_r(x^\star))}{M_r} \geqslant 1 \\
&\iff \limsup_{r \to 0} \frac{M_r}{\mu(B_r(x^\star))} \leqslant 1.
\end{aligned}
$$

This motivates another definition:

Definition 2 (after Helin and Burger (2015))

A **global weak mode** of $\mu \in \mathcal{P}(X)$ is any $x^\star \in X$ such that

$$\limsup_{r \to 0} \frac{\mu(B_r(x'))}{\mu(B_r(x^\star))} \leqslant 1 \text{ for all } x' \in X.$$

Defining a mode of a measure. 3: Minimisers of OM functionals

Definition 3

An **Onsager–Machlup (OM) functional** for $\mu \in \mathcal{P}(X)$ is a function $I_\mu \colon E \to \mathbb{R}$ with

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

We call $E \subseteq X$ the **domain** of the OM functional.

- OM functionals are at most unique up to addition of constants — this aspect requires some care, which this presentation will neglect!
- If $\mu \in \mathcal{P}(\mathbb{R}^d)$ has Lebesgue density ρ, then $I_\mu := -\log \rho$ is an OM functional for μ.
- Any measure admits an OM functional if E is small enough.
- Can measures on "large" spaces have OM functionals with large E?
- Are minimisers of I_μ "most probable points" under μ? (Cf. Dürr and Bach (1978).)

The Gaussian OM functional

- The prime example of an OM functional is the OM functional of a centred Gaussian measure $\mu = \mathcal{N}(0, C)$ on a separable Hilbert space X.
- Here, for simplicity, we assume that the covariance operator $C \colon X \to X$,

$$\langle u, Cv \rangle := \int_X \langle u, x \rangle \langle v, x \rangle \, \mu(\mathrm{d}x)$$

which is always symmetric and positive semi-definite, is actually positive definite.
- In this case, μ has the OM functional $I_\mu \colon H(\mu) := \operatorname{ran} C^{1/2} \to \mathbb{R}$

$$I_\mu(u) = \frac{1}{2} \| C^{-1/2} u \|^2 \text{ for } u \in H(\mu).$$

- Furthermore, one can show that, for $u \notin H(\mu)$, $\lim_{r \to 0} \frac{\mu(B_r(u))}{\mu(B_r(0))} = 0$, so we can sensibly think of I_μ as taking the value $+\infty$ there.

A new-ish formal property

We formalise a property used implicitly in e.g. Dashti et al. (2013):

Definition 4

We will say that **property** $M(\mu, E)$ holds for $\mu \in \mathcal{P}(X)$ and $E \subseteq X$ if, for some $x^\star \in E$,

$$x \in X \setminus E \implies \lim_{r \to 0} \frac{\mu(B_r(x))}{\mu(B_r(x^\star))} = 0.$$

- Property $M(\mu, E)$ always holds if E is large enough.
- Property $M(\mu, E)$ does not say that $\mu(X \setminus E) = 0$!
- Property $M(\mu, E)$ does say that points outside E cannot qualify as modes of μ.
- Standard example: a Gaussian measure $\mu = \mathcal{N}(0, C)$ on an infinite-dimensional Hilbert space X with infinite-dimensional **Cameron–Martin space** $H(\mu) :=$ **ran** $C^{1/2}$ satisfies property $M(\mu, H(\mu))$ and yet has $\mu(H(\mu)) = 0$.

OM functionals, property M, and modes

Lemma 5 (Ayanbayev et al., 2022a, Prop. 4.1)

Let μ have OM functional $I_\mu \colon E \to \mathbb{R}$ and satisfy property $M(\mu, E)$. Set $I_\mu(x) := +\infty$ for $x \notin E$. Then the global weak modes of μ are precisely the global minimisers of $I_\mu \colon X \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$.

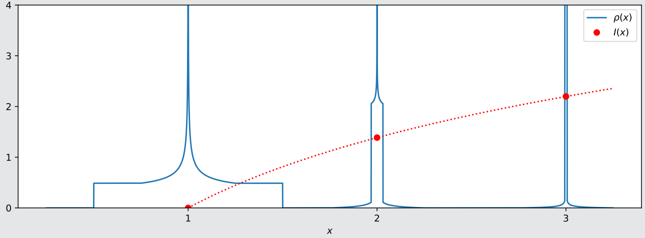
- This result gives a rigorous meaning to the claim of Dürr and Bach (1978) that OM minimisers should be seen as "most likely points" in the sense of global weak modes.
- Unfortunately, it is **not** generally true that strong modes are OM-minimisers, even when such minimisers exist and property M holds!

OM functionals, property M, and modes

Example 6

Let $\mu \in \mathcal{P}(\mathbb{R})$ have Lebesgue density $\rho := \frac{24}{5\pi^2} \sum_{k \in \mathbb{N}} \rho_k$, where

$$\rho_0(x) := \tfrac{1}{4}\left(|x|^{-1/2} - 2\right) \mathbb{1}_{[-\frac{1}{4}, \frac{1}{4}] \setminus \{0\}}(x), \qquad \rho_k(x) := \frac{\rho_0(x-k)}{k^2} + k^2\, \mathbb{1}_{\left[-\frac{1}{2k^4}, \frac{1}{2k^4}\right]}(x-k).$$



The measure μ has OM functional $I_\mu(x) = 2 \log x$ for $x \in E = \mathbb{N}$, this domain cannot be extended, and property $M(\mu, \mathbb{N})$ holds. However, μ has a global weak mode at 1, and this minimises I_μ, but it is not a strong mode.

Γ-convergence: A capsule summary

- Γ-convergence, originating with De Giorgi and Franzoni (1975), is a principal example of a kind of variational convergence for functionals $F_n \colon X \to \overline{\mathbb{R}}$.
- The idea, under suitable assumptions, is to have a notion of convergence for functionals so that
$$F_n \xrightarrow[n\to\infty]{\Gamma} F \implies \operatorname*{arg\,min}_X F_n \xrightarrow[n\to\infty]{} \operatorname*{arg\,min}_X F.$$
i.e. the minimisers of F_n converge to the minimisers of F.
- Γ-convergence, and related notions such as Mosco convergence, have met great success in the study of optimisation problems in general and the calculus of variations in particular.

Definition 7

Given extended-real-valued functions $F_n, F\colon X \to \overline{\mathbb{R}}$, we say that F_n **Γ-converges** to F, written $F_n \xrightarrow[n\to\infty]{\Gamma} F$, if, for every $x \in X$,

- (Γ-lim inf inequality) for every sequence $(x_n)_{n\in\mathbb{N}}$ converging to x,

$$F(x) \leqslant \liminf_{n\to\infty} F_n(x_n);$$

- (Γ-lim sup inequality) and there exists a "recovery sequence" $(x_n)_{n\in\mathbb{N}}$ converging to x such that

$$F(x) \geqslant \limsup_{n\to\infty} F_n(x_n).$$

Pointwise, continuous, uniform and Γ-convergence

Definition 8

For $F_n, F \colon X \to \overline{\mathbb{R}}$ as before, we say that F_n **converges continuously** to F if, for every $x \in X$ and every neighbourhood V of $F(x)$ in $\overline{\mathbb{R}}$, there exists $N \in \mathbb{N}$ and $r > 0$ such that

$$(n \geqslant N \text{ and } d(x', x) < r) \implies F_n(x') \in V.$$

$$\begin{tikzcd}
F_n \xrightarrow[n\to\infty]{\text{unif}} F \arrow[r, Rightarrow, "\text{if } F \text{ continuous}"] & F_n \xrightarrow[n\to\infty]{\text{cts}} F \arrow[r, Rightarrow] \arrow[dr, Rightarrow] & F_n \xrightarrow[n\to\infty]{\Gamma} F \arrow[d, Rightarrow, shift right, "/" marking] \\
& & F_n \xrightarrow[n\to\infty]{\text{pt}} F \arrow[u, Rightarrow, shift right, "/" marking]
\end{tikzcd}$$

$$\left(F_n \xrightarrow[n\to\infty]{\Gamma} F \text{ and } F_n \xrightarrow[n\to\infty]{\text{pt}} G \right) \implies F \leqslant G.$$

$$\left(F_n \xrightarrow[n\to\infty]{\text{cts}} F \text{ and } G_n \xrightarrow[n\to\infty]{\Gamma} G \right) \implies F_n + G_n \xrightarrow[n\to\infty]{\Gamma} F + G.$$

Definition 9

We say that $(F_n)_{n\in\mathbb{N}}$ is **equicoercive** if for all $t \in \mathbb{R}$, there exists a compact $K_t \subseteq X$ such that, for all $n \in \mathbb{N}$, $F_n^{-1}([-\infty, t]) \subseteq K_t$.

The fundamental theorem of Γ-convergence

Definition 9

We say that $(F_n)_{n\in\mathbb{N}}$ is **equicoercive** if for all $t \in \mathbb{R}$, there exists a compact $K_t \subseteq X$ such that, for all $n \in \mathbb{N}$, $F_n^{-1}([-\infty, t]) \subseteq K_t$.

Theorem 10 (Fundamental theorem of Γ-convergence; Braides, 2006, Theorem 2.10)

Suppose that $F_n, F \colon X \to \overline{\mathbb{R}}$ are such that $F_n \xrightarrow[n\to\infty]{\Gamma} F$ and $(F_n)_{n\in\mathbb{N}}$ is equicoercive. Then

- *F has a minimum value and $\min_X F = \lim_{n\to\infty} \inf_X F_n$;*
- *if $(x_n)_{n\in\mathbb{N}}$ is a precompact sequence such that $\lim_{n\to\infty} F_n(x_n) = \min_X F$, then every limit of a convergent subsequence of $(x_n)_{n\in\mathbb{N}}$ is a minimiser of F; and*
- *if each F_n has a minimiser $x_n^\star$, then every convergent subsequence of $(x_n^\star)_{n\in\mathbb{N}}$ has as its limit a minimiser of F.*

(The hypotheses of the fundamental theorem can be relaxed somewhat to use only "equi-mild coercivity".)

Γ-convergence of OM functionals

Γ-convergence of OM functionals and convergence of modes

We're in a position to state our first theorem, and it comes almost for free...

Theorem 11 (Γ-convergence and equicoercivity imply convergence of modes; Ayanbayev et al. (2022a, Theorem 4.2))

For $n \in \mathbb{N} \cup \{\infty\}$, let $\mu^{(n)} \in \mathcal{P}(X)$ have OM functional $I_{\mu^{(n)}} \colon E^{(n)} \to \mathbb{R}$ and satisfy property $M(\mu^{(n)}, E^{(n)})$; extend each $I_{\mu^{(n)}}$ to take the value $+\infty$ on $X \setminus E^{(n)}$. Suppose that the sequence $(I_{\mu^{(n)}})_{n \in \mathbb{N}}$ is equicoercive and Γ-converges to $I_{\mu^{(\infty)}}$. Then, if $u^{(n)}$ is a global weak mode of $\mu^{(n)}$, $n \in \mathbb{N}$, every convergent subsequence of $(u^{(n)})_{n \in \mathbb{N}}$ has as its limit a global weak mode of $\mu^{(\infty)}$.

Γ-convergence of OM functionals and convergence of modes

We're in a position to state our first theorem, and it comes almost for free...

Theorem 11 (Γ-convergence and equicoercivity imply convergence of modes; Ayanbayev et al. (2022a, Theorem 4.2))

For $n \in \mathbb{N} \cup \{\infty\}$, let $\mu^{(n)} \in \mathcal{P}(X)$ have OM functional $I_{\mu^{(n)}} \colon E^{(n)} \to \mathbb{R}$ and satisfy property $M(\mu^{(n)}, E^{(n)})$; extend each $I_{\mu^{(n)}}$ to take the value $+\infty$ on $X \setminus E^{(n)}$. Suppose that the sequence $(I_{\mu^{(n)}})_{n \in \mathbb{N}}$ is equicoercive and Γ-converges to $I_{\mu^{(\infty)}}$. Then, if $u^{(n)}$ is a global weak mode of $\mu^{(n)}$, $n \in \mathbb{N}$, every convergent subsequence of $(u^{(n)})_{n \in \mathbb{N}}$ has as its limit a global weak mode of $\mu^{(\infty)}$.

Proof.

The global weak modes are exactly the minimisers of the extended OM functionals, and the rest follows from the fundamental theorem of Γ-convergence. □

Γ-convergence of OM functionals and convergence of modes

- The pathological examples of non-convergent modes given earlier fall outside the realm of Theorem 11: the negative log-densities involved converge pointwise but not uniformly, and indeed do not Γ-converge.
- Theorem 11 is very general, and there is no free lunch: one does need to verify equicoercivity and Γ-convergence for the application at hand.
- Let's examine the Γ-convergence and equicoercivity of the OM functionals of measures that are often used as priors in BIPs, even though their modes are quite obvious.
- This only *looks* like a trivial exercise: Γ-convergence and equicoercivity of posterior OM functionals — i.e. for reweightings of these priors — and hence convergence of MAP estimators, will follow later.

Skip discussion of pseudoinverse square roots ▶

A digression on pseudoinverses and pseudoinverse square roots

- "Everyone knows" that the OM functional of a Gaussian measure is one half the square of its Cameron–Martin norm.
- To make this statement precise, we need to be precise about the inverse square root of its (possibly indefinite) covariance operator.

Definition 12

For a bounded linear operator A between Hilbert spaces X and Y, the **Moore–Penrose pseudoinverse** $A^\dagger$ of A is the unique extension of $(A|_{(\ker A)^\perp})^{-1}$ to a (generally unbounded) linear operator $A^\dagger \colon \operatorname{ran} A \oplus (\operatorname{ran} A)^\perp \to X$ subject to the restriction that $\ker A^\dagger = (\operatorname{ran} A)^\perp$.

For $y \in \operatorname{ran} A \oplus (\operatorname{ran} A)^\perp$,

$$A^\dagger y = \arg\min \big\{ \|x\|_X \,\big|\, x \text{ minimises } \|Ax - y\| \big\}.$$

In particular, for $y \in \operatorname{ran} A$, $A^\dagger y$ is the minimum-norm solution of $Ax = y$.

A digression on pseudoinverses and pseudoinverse square roots

Definition 13

For a compact SPSD operator $C = \sum_{n\in\mathbb{N}} \sigma_n^2 \, e_n \otimes e_n$ on a Hilbert space X, $(e_n)_{n\in\mathbb{N}}$ being an orthonormal system in X and $\sigma_n \geqslant 0$ for each $n \in \mathbb{N}$, we denote the SPSD operator square root of C by $C^{1/2}$ and furthermore set

$$C^{\dagger/2} := (C^{1/2})^{\dagger} = \sum_{n\in\mathbb{N}\,:\,\sigma_n\neq 0} \sigma_n^{-1} \, e_n \otimes e_n.$$

Note that $(C^{\dagger})^{1/2}$ can differ from $(C^{1/2})^{\dagger}$ since it may have a smaller domain.

OM functionals for Gaussian measures

Lemma 14 (Ayanbayev et al., 2022a, Cor. 5.4)

The extended OM functional of $\mu = \mathcal{N}(m, C)$ on a separable Hilbert space X is $I_\mu \colon X \to \overline{\mathbb{R}}$,

$$I_\mu(u) := \begin{cases} \frac{1}{2}\left\| C^{\dagger/2}(u-m) \right\|_X^2 & \text{for } u - m \in H(\mu) = \operatorname{ran} C^{1/2}, \\ +\infty & \text{otherwise,} \end{cases}$$

and property $M(\mu, m + H(\mu))$ holds.

OM functionals for Gaussian measures

Lemma 14 (Ayanbayev et al., 2022a, Cor. 5.4)

The extended OM functional of $\mu = \mathcal{N}(m, C)$ *on a separable Hilbert space* X *is* $I_\mu \colon X \to \overline{\mathbb{R}}$,

$$I_\mu(u) := \begin{cases} \frac{1}{2} \left\| C^{\dagger/2}(u - m) \right\|_X^2 & \text{for } u - m \in H(\mu) = \operatorname{ran} C^{1/2}, \\ +\infty & \text{otherwise,} \end{cases}$$

and property $M(\mu, m + H(\mu))$ *holds.*

Theorem 15 (Γ-convergence and equicoercivity of Gaussian OM functionals; Ayanbayev et al., 2022a, Thm. 5.5)

Let X *be a separable Hilbert space and* $\mu^{(n)} = \mathcal{N}(m^{(n)}, C^{(n)})$, *for* $n \in \mathbb{N} \cup \{\infty\}$, *be Gaussian measures on* X. *Then*

$$\left. \begin{array}{c} \left\| m^{(n)} - m^{(\infty)} \right\|_X \to 0 \text{ and} \\ \left\| C^{(n)} - C^{(\infty)} \right\|_{\mathrm{op}} \to 0 \end{array} \right\} \implies \left\{ \begin{array}{c} I_{\mu^{(n)}} \xrightarrow[n \to \infty]{\Gamma} I_{\mu^{(\infty)}} \text{ and} \\ (I_{\mu^{(n)}})_{n \in \mathbb{N}} \text{ is equicoercive.} \end{array} \right.$$

Besov measures

- **Besov priors** (Lassas et al., 2009; Dashti et al., 2012; Agapiou et al., 2018) have been advocated as an extension of Gaussian priors for BIPs.
- Besov priors have two key parameters: **"smoothness"** $s \in \mathbb{R}$ and **"integrability"** $p \geqslant 1$; for historical reasons to do with connections to PDE theory, there is also a "spatial dimension" $d \in \mathbb{N}$ and the quantity $^{s}/_{d}$ occurs often.
- The case $p = 2$ corresponds to Gaussian distributions.
- The case $p = 1$ has been studied for its sparsifying / edge-preserving properties (contrast with TV regularisation, Lassas and Siltanen (2004)).
- Just to keep the notation somewhat under control, this talk will concentrate on the case $p = 1$ and study stability w.r.t. smoothness s, but our results do cover general p and a large class of more general product priors and their perturbations (Ayanbayev et al., 2022b).

Besov-1 measures

- Let $s \in \mathbb{R}$, $d \in \mathbb{N}$, $\eta > 0$, $t := s - d(1 + \eta)$.
- The parameter s is thought of as a "smoothness parameter" and d as a "spatial dimension". The parameter t is "a bit less smooth" than s.
- Define $\gamma_0 := 1$ and $\gamma, \delta \in \mathbb{R}^{\mathbb{N}}$ by

$$\gamma_k := k^{1-s/d-1/2}, \qquad \delta_k := k^{1-t/d-1/2} = k^{2+\eta-s/d-1/2}, \qquad k \in \mathbb{N},$$

  and let $\mu_k \in \mathcal{P}(\mathbb{R})$ for $k \in \mathbb{N} \cup \{0\}$ have the Lebesgue density

$$\frac{\mathrm{d}\mu_k}{\mathrm{d}u}(u) = \frac{1}{2\gamma_k^{-1}} \exp(-|u/\gamma_k|).$$

Definition 16 (Sequence space Besov measures and Besov spaces)

We call $\mu := \bigotimes_{k\in\mathbb{N}} \mu_k$ a (**sequence space**) **Besov measure** on $\mathbb{R}^{\mathbb{N}}$ and write $B_1^s := \mu$. The corresponding **Besov space** is the weighted sequence space $(X_1^s, \|\cdot\|_{X_1^s}) := (\ell_\gamma^1, \|\cdot\|_{\ell_\gamma^1})$, i.e.

$$\|h\|_{X_1^s} := \sum_{k\in\mathbb{N}} k^{s/d-1/2}|h_k|$$

Definition 16 (Sequence space Besov measures and Besov spaces)

We call $\mu := \bigotimes_{k\in\mathbb{N}} \mu_k$ a (**sequence space**) **Besov measure** on $\mathbb{R}^{\mathbb{N}}$ and write $B_1^s := \mu$. The corresponding **Besov space** is the weighted sequence space $(X_1^s, \|\cdot\|_{X_1^s}) := (\ell_\gamma^1, \|\cdot\|_{\ell_\gamma^1})$, i.e.

$$\|h\|_{X_1^s} := \sum_{k\in\mathbb{N}} k^{s/d-1/2}|h_k|$$

- One can perform the same construction in any separable Hilbert space instead of $\ell^2 \subset \mathbb{R}^{\mathbb{N}}$, considering random expansion w.r.t. a countable complete orthonormal basis.
- In the case of $L^2(\mathbb{T}^d;\mathbb{R})$ with the Fourier basis, X_1^s is the Besov space B_{11}^s (hence the name).

OM functionals for Besov-1 measures

- One thinks of B_1^s as having a formal Lebesgue density proportional to $\exp(-\|\cdot\|_{X_1^s})$ in the same way that $\mathcal{N}(0, C)$ has a formal density proportional to $\exp(-\frac{1}{2}\|C^{-1/2}\cdot\|^2)$.
- But is this actually true on the level of OM functionals?

OM functionals for Besov-1 measures

- One thinks of B_1^s as having a formal Lebesgue density proportional to $\exp(-\|\cdot\|_{X_1^s})$ in the same way that $\mathcal{N}(0, C)$ has a formal density proportional to $\exp(-\frac{1}{2}\|C^{-1/2}\cdot\|^2)$.
- But is this actually true on the level of OM functionals?

Lemma 17 (Support of a Besov-1 measure; Ayanbayev et al., 2022a, Lem. 5.10)

Let $\mu = B_1^s$ be the Besov measure defined above and $X = X_1^t = \ell_\delta^1$. Then $\mu(X) = 1$.

Proposition 18 (OM functional of a Besov-1 measure; Ayanbayev et al., 2022a, Prop. 5.11)

Let $\mu = B_1^s$ on the space $X = X_1^t = \ell_\delta^1$. Then property $M(\mu, X_1^s)$ is satisfied and the OM functional $I_\mu \colon X_1^t \to \overline{\mathbb{R}}$ of μ is given by

$$I_\mu(u) = \begin{cases} \|u\|_{X_1^s} & \text{for } u \in X_1^s, \\ \infty & \text{otherwise.} \end{cases}$$

OM functionals for Besov-1 measures

Theorem 19 (Γ-convergence and equicoercivity of Besov-1 OM functionals; Ayanbayev et al., 2022a, Thm. 5.13)

Let $\mu^{(n)} := B_1^{s^{(n)}}$, $n \in \mathbb{N} \cup \{+\infty\}$, be centered Besov measures such that $s^{(n)} \to s^{(\infty)}$. Then there exists $n_0 \in \mathbb{N}$ such that, for each $n \geqslant n_0$, $\mu^{(n)}(\ell^1_{\delta(\infty)}) = 1$ and we therefore consider these measures on $X = X_1^{t^{(\infty)}} = \ell^1_{\delta(\infty)}$ (after dropping the first $n_0 - 1$ measures). Then the associated OM functionals $I_{\mu^{(n)}} = \| \cdot \|_{X_1^{s^{(n)}}} : X \to \overline{\mathbb{R}}$, $n \geqslant n_0$, are equicoercive and

$$I_{\mu^{(n)}} \xrightarrow[n\to\infty]{\Gamma} I_{\mu^{(\infty)}}.$$

- For emphasis: each of the measures $B_1^{s^{(n)}}$ is centred, with the origin being both the mean and the mode. Convergence of modes is therefore trivial.
- However, Γ-convergence of the OM functionals is not trivial — it is essential for the study of Γ-convergence of posterior OM functionals in the next step.

A sketch of some generalisations

- Besov-p measures with $1 \leqslant p \leqslant 2$ and mean $m \in X$
  - Infinite product of marginal densities $\propto \exp(-|\frac{u_k - m_k}{\gamma_k}|^p)$
  - OM functional is $\|u - m\|_{X_p^s}^p$ on $m + X_p^t$, with property $M(\mu, m + X_p^s)$. ✓
  - Γ-convergence and equicoercivity with respect to mean and smoothness. ✓
- Cauchy measures
  - Countable products of marginal densities $\propto \big(1 + |\frac{u_k - m_k}{\gamma_k}|\big)^{-1}$.
  - OM functional is $\sum_k \log(1 + \gamma_k^{-2}(u_k - m_k))$ with property $M(\mu, m + \ell_\gamma^2)$. ✓
  - Γ-convergence and equicoercivity with respect to location and scale parameters. ✓
- General scaled product measures
  - Countable products of marginal densities $\rho_k(u_u) \propto \rho_0\big(\frac{u_k - m_k}{\gamma_k}\big)$, with ρ_0 a "nice" reference density on $\mathbb{R}$
  - OM functional is more or less what it should be (lower bound is relatively straightforward, upper bound only in some cases, maximal domain and property M are also tricky. . .) $\approx$ ✓
  - Γ-convergence and equicoercivity with respect to location and scale parameters. ✓

Bayesian inverse problems

Bayesian inverse problems

- An inverse problem consists of the recovery of an unknown u from related observational data y. In the Bayesian approach to inverse problems (Kaipio and Somersalo, 2005; Stuart, 2010), these two objects are treated as coupled random variables $\boldsymbol{u}$ and $\boldsymbol{y}$ that take values in spaces X and Y respectively.
- A priori knowledge about $\boldsymbol{u}$ is represented by a prior probability measure $\mu_0 \in \mathcal{P}(X)$ and one is given access to a realisation y of $\boldsymbol{y}$. One also posits a likelihood model $\ell \colon X \to \mathcal{P}(Y)$.
- The solution of the BIP is, by definition, the posterior probability measure $\mu^y \in \mathcal{P}(X)$, i.e. the conditional distribution of $\boldsymbol{u}$ given that $\boldsymbol{y} = y$, or the disintegration of the joint distribution $\mu(\mathrm{d}u, \mathrm{d}y) \propto \mu_0(\mathrm{d}u)\ell(\mathrm{d}y|u)$ of $(\boldsymbol{u}, \boldsymbol{y})$ along the y-fibre (Chang and Pollard, 1997).

Bayesian inverse problems

- For simplicity, focus on the case that μ^y has a density with respect to μ_0 of the form

$$\mu^y(\mathrm{d}u) \propto \exp(-\Phi(u; y))\, \mu_0(\mathrm{d}u).$$

- The potential $\Phi\colon X \times Y \to \mathbb{R}$ encodes both the idealised relationship between the unknown and the data and statistical assumptions about any observational noise.
- Textbook example: X is a separable Hilbert or Banach space of functions, $Y = \mathbb{R}^J$ for some $J \in \mathbb{N}$, and that $\boldsymbol{y} = \mathcal{O}(\boldsymbol{u}) + \boldsymbol{\eta}$ for some deterministic observation map $\mathcal{O}\colon X \to Y$ and additive non-degenerate Gaussian noise $\boldsymbol{\eta} \sim \mathcal{N}(0, C_{\boldsymbol{\eta}})$ that is a priori independent of $\boldsymbol{u}$, in which case Φ is the familiar quadratic misfit

$$\Phi(u; y) = \frac{1}{2} \big\| C_{\boldsymbol{\eta}}^{-1/2} (y - \mathcal{O}(u)) \big\|^2.$$

Bayesian inverse problems

- So, on a hand-wavy level, the posterior $\mu^y(\mathrm{d}u) \propto \exp(-\Phi(u; y))\, \mu_0(\mathrm{d}u)$ has a "negative log-Lebesgue density"

$$-\log \rho^y(u) = \underbrace{\Phi(u; y)}_{\text{misfit}} \underbrace{- \log \rho_0(u)}_{\text{regularisation}}.$$

- In the case of a Gaussian prior $\mu_0 = \mathcal{N}(m_0, C_0)$, this is Tikhonov–Philips regularisation:

$$-\log \rho^y(u) = \Phi(u; y) + \frac{1}{2}\langle u, C_0^{-1} u\rangle.$$

- This is the connection between the Bayesian viewpoint and the regularised optimisation viewpoint on inverse problems:
  Minimisers of the posterior "negative log-Lebesgue density" ought to be regarded as "most probable points for μ^y".

MAP estimation for BIPs

Consequences for MAP estimation in BIPs

- In view of the earlier discussion, we can be more rigorous in our statements about MAP estimators.
- We want to be able to define "MAP estimator" to mean "global weak mode of the posterior"...

Consequences for MAP estimation in BIPs

- In view of the earlier discussion, we can be more rigorous in our statements about MAP estimators.
- We want to be able to define "MAP estimator" to mean "global weak mode of the posterior"...

  ...and say that these points are minimisers of the OM functional of the posterior...

Consequences for MAP estimation in BIPs

- In view of the earlier discussion, we can be more rigorous in our statements about MAP estimators.
- We want to be able to define "MAP estimator" to mean "global weak mode of the posterior"...

  ...and say that these points are minimisers of the OM functional of the posterior...

  ...that the OM functional of the posterior is Φ plus the OM functional of the prior...

Consequences for MAP estimation in BIPs

- In view of the earlier discussion, we can be more rigorous in our statements about MAP estimators.
- We want to be able to define "MAP estimator" to mean "global weak mode of the posterior"...

  ...and say that these points are minimisers of the OM functional of the posterior...

  ...that the OM functional of the posterior is Φ plus the OM functional of the prior...

  ...and that the MAP estimators are stable under suitable continuous convergence / Γ-convergence / equicoercivity assumptions on Φ and I_{μ_0}.

Consequences for MAP estimation in BIPs

- In view of the earlier discussion, we can be more rigorous in our statements about MAP estimators.
- We want to be able to define "MAP estimator" to mean "global weak mode of the posterior"...
  ...and say that these points are minimisers of the OM functional of the posterior...
  ...that the OM functional of the posterior is Φ plus the OM functional of the prior...
  ...and that the MAP estimators are stable under suitable continuous convergence / Γ-convergence / equicoercivity assumptions on Φ and I_{μ_0}.
- And this is indeed what we can show!

Skip the formal statement of the theorem ▶

Transfer of property M, Γ-convergence, equicoercivity, etc.

Theorem 20 (Ayanbayev et al. (2022a, Theorem 6.1))

For each $n \in \mathbb{N} \cup \{\infty\}$, let $\mu_0^{(n)} \in \mathcal{P}(X)$ and let $\Phi^{(n)} \colon X \to \mathbb{R}$ be locally uniformly continuous. Suppose that, for each $n \in \mathbb{N} \cup \{\infty\}$

$$\mu^{(n)}(\mathrm{d}x) := \frac{1}{Z^{(n)}} e^{-\Phi^{(n)}(x)}\, \mu_0^{(n)}(\mathrm{d}x), \qquad Z^{(n)} := \int_X e^{-\Phi^{(n)}(x)}\, \mu_0^{(n)}(\mathrm{d}x) \in (0, \infty),$$

and each $\mu_0^{(n)}$ has an OM functional $I_{\mu_0^{(n)}} \colon E^{(n)} \to \mathbb{R}$. Then:

1. *Each $\mu^{(n)}$ has $I_{\mu^{(n)}} := \Phi^{(n)} + I_{\mu_0^{(n)}} \colon E^{(n)} \to \mathbb{R}$ as an OM functional.*
2. *Suppose that property $M(\mu_0^{(n)}, E^{(n)})$ holds. Then property $M(\mu^{(n)}, E^{(n)})$ also holds, and the global weak modes of $\mu_0^{(n)}$ (resp. of $\mu^{(n)}$) are the global minimisers of the extended OM functional $I_{\mu_0^{(n)}} \colon X \to \overline{\mathbb{R}}$ (resp. of $I_{\mu^{(n)}} \colon X \to \overline{\mathbb{R}}$).*
3. *If $I_{\mu_0^{(n)}} \xrightarrow[n\to\infty]{\Gamma} I_{\mu_0^{(\infty)}}$ and $\Phi^{(n)} \xrightarrow[n\to\infty]{\text{cts}} \Phi^{(\infty)}$ as $n \to \infty$, then $I_{\mu^{(n)}} \xrightarrow[n\to\infty]{\Gamma} I_{\mu^{(\infty)}}$.*

Transfer of property M, Γ-convergence, equicoercivity, etc.

Theorem 20 (Ayanbayev et al. (2022a, Theorem 6.1))

For each $n \in \mathbb{N} \cup \{\infty\}$, let $\mu_0^{(n)} \in \mathcal{P}(X)$ and let $\Phi^{(n)} \colon X \to \mathbb{R}$ be locally uniformly continuous. Suppose that, for each $n \in \mathbb{N} \cup \{\infty\}$

$$\mu^{(n)}(\mathrm{d}x) := \frac{1}{Z^{(n)}} e^{-\Phi^{(n)}(x)} \, \mu_0^{(n)}(\mathrm{d}x), \qquad Z^{(n)} := \int_X e^{-\Phi^{(n)}(x)} \, \mu_0^{(n)}(\mathrm{d}x) \in (0, \infty),$$

and each $\mu_0^{(n)}$ has an OM functional $I_{\mu_0^{(n)}} \colon E^{(n)} \to \mathbb{R}$. Then:

4. *If $(I_{\mu_0^{(n)}})_{n \in \mathbb{N}}$ is equicoercive and the functions $\Phi^{(n)}$ are uniformly bounded from below by some constant $M \in \mathbb{R}$, then $(I_{\mu^{(n)}})_{n \in \mathbb{N}}$ is also equicoercive with respect to the same representatives of $I_{\mu^{(n)}}$ as for the Γ-convergence.*
5. *Under the assumptions of parts 2–4, the cluster points as $n \to \infty$ of the global weak modes of the posteriors $\mu^{(n)}$ are the global weak modes of the limiting posterior $\mu^{(\infty)}$.*

Consequences for BIPs

Consider a BIP with prior μ_0, potential Φ bounded below, and observed data y, each of which may now be approximated. In addition to the assumptions of Theorem 20, assume for simplicity that I_{μ_0} is lower semicontinuous, so that it equals its own Γ-limit.

- If the potential Φ and prior μ_0 are held constant and we examine the posterior $\mu^{(n)}$ associated to data $y^{(n)}$, then

$$\Phi(\cdot\,; y^{(n)}) \xrightarrow[n\to\infty]{\text{cts}} \Phi(\cdot\,; y) \implies \begin{cases} I_{\mu^{(n)}} \xrightarrow[n\to\infty]{\Gamma} I_{\mu} \text{ and} \\ (I_{\mu^{(n)}})_{n\in\mathbb{N}} \text{ is equicoercive} \end{cases}$$

$$\implies \text{convergence of MAP estimators (up to subsequences)}$$

Consequences for BIPs

Consider a BIP with prior μ_0, potential Φ bounded below, and observed data y, each of which may now be approximated. In addition to the assumptions of Theorem 20, assume for simplicity that I_{μ_0} is lower semicontinuous, so that it equals its own Γ-limit.

- If the data y and potential Φ are held constant and we examine the posterior $\mu^{(n)}$ associated to prior $\mu^{(n)}$, then

$$\left.\begin{array}{c} I_{\mu_0^{(n)}} \xrightarrow[n\to\infty]{\Gamma} I_{\mu_0} \text{ and} \\ (I_{\mu_0^{(n)}})_{n\in\mathbb{N}} \text{ is equicoercive} \end{array}\right\} \implies \left\{\begin{array}{c} I_{\mu^{(n)}} \xrightarrow[n\to\infty]{\Gamma} I_{\mu} \text{ and} \\ (I_{\mu^{(n)}})_{n\in\mathbb{N}} \text{ is equicoercive} \end{array}\right.$$

$$\implies \text{convergence of MAP estimators (up to subsequences)}$$

Consequences for BIPs

Consider a BIP with prior μ_0, potential Φ bounded below, and observed data y, each of which may now be approximated. In addition to the assumptions of Theorem 20, assume for simplicity that I_{μ_0} is lower semicontinuous, so that it equals its own Γ-limit.

- Finally, if the data and prior are held constant and we examine the posterior $\mu^{(n)}$ associated to the potential $\Phi^{(n)}$, then

$$\Phi^{(n)}(\,\cdot\,; y) \xrightarrow[n\to\infty]{\text{cts}} \Phi(\,\cdot\,; y) \implies \begin{cases} I_{\mu^{(n)}} \xrightarrow[n\to\infty]{\Gamma} I_{\mu} \text{ and} \\ (I_{\mu^{(n)}})_{n\in\mathbb{N}} \text{ is equicoercive} \end{cases}$$

$$\implies \text{convergence of MAP estimators (up to subsequences)}$$

In particular, this holds when the approximate misfit/potential $\Phi^{(n)}$ arises through projection, e.g. Galerkin discretisation.

Closing remarks

Closing remarks

- We have established a stability theory for non-parametric MAP estimators by focussing on global weak modes, which are characterised as minimisers of extended Onsager–Machlup functionals, and then studying the variational Γ-convergence of these functionals.
- Our analysis encompasses Bayesian posteriors associated to Gaussian, Besov, and Cauchy priors and reveals simple sufficient conditions for stability of MAP estimators (continuous convergence of log-likelihoods, Γ-convergence and equicoercivity of prior OM functionals).
- These conditions could be added to the now-standard conditions for stability of the BIP à la Stuart (2010) to ensure stability of *both* the BIP and the MAP estimation problem. (There are hypotheses that imply both BIP stability and MAP stability, but the BIP and MAP stability assumptions are generally independent.)
- Open problems / avenues for further work:
    - Unfortunately Γ-convergence + equicoercivity alone cannot deliver a convergence rate for the modes!
    - Other classes of priors, e.g. hierarchical and deep priors, priors on non-linear spaces such as shape spaces, etc.

Thank You!

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 Probl.*, 34(4):045002, 37, 2018. doi:10.1088/1361-6420/aaacac.

B. Ayanbayev, I. Klebanov, H. C. Lie, and T. J. Sullivan. Γ-convergence of Onsager–Machlup functionals: I. With applications to maximum a posteriori estimation in Bayesian inverse problems. *Inverse Probl.*, 38(2):025005, 32pp., 2022a. doi:10.1088/1361-6420/ac3f81.

B. Ayanbayev, I. Klebanov, H. C. Lie, and T. J. Sullivan. Γ-convergence of Onsager–Machlup functionals: II. Infinite product measures on Banach spaces. *Inverse Probl.*, 38(2):025006, 35pp., 2022b. doi:10.1088/1361-6420/ac3f82.

A. Braides. A handbook of Γ-convergence. In *Handbook of Differential Equations: Stationary Partial Differential Equations*, volume 3, pages 101–213. 2006. doi:10.1016/S1874-5733(06)80006-9.

J. T. Chang and D. Pollard. Conditioning as disintegration. *Statist. Neerlandica*, 51(3):287–317, 1997. doi:10.1111/1467-9574.00056.

C. Clason, T. Helin, R. Kretschmann, and P. Piiroinen. Generalized modes in Bayesian inverse problems. *SIAM/ASA J. Uncertain. Quantif.*, 7(2):652–684, 2019. doi:10.1137/18M1191804.

M. Dashti, S. Harris, and A. Stuart. Besov priors for Bayesian inverse problems. *Inverse Probl. Imaging*, 6(2):183–200, 2012. doi:10.3934/ipi.2012.6.183.

M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss. MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Probl.*, 29(9):095017, 27, 2013. doi:10.1088/0266-5611/29/9/095017.

References 2

E. De Giorgi and T. Franzoni. Su un tipo di convergenza variazionale. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat. (8)*, 58(6):842–850, 1975. ISSN 0392-7881.

D. Dürr and A. Bach. The Onsager–Machlup function as Lagrangian for the most probable path of a diffusion process. *Comm. Math. Phys.*, 60(2):153–170, 1978. doi:10.1007/BF01609446.

T. Helin and M. Burger. Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Probl.*, 31(8):085009, 22, 2015. doi:10.1088/0266-5611/31/8/085009.

J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*, volume 160 of *Applied Mathematical Sciences*. Springer, New York, 2005. doi:10.1007/b138659.

M. Lassas and S. Siltanen. Can one use total variation prior for edge-preserving Bayesian inversion? *Inverse Probl.*, 20(5): 1537–1563, 2004. doi:10.1088/0266-5611/20/5/013.

M. Lassas, E. Saksman, and S. Siltanen. Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Probl. Imaging*, 3(1):87–122, 2009. doi:10.3934/ipi.2009.3.87.

B. Sprungk. On the local Lipschitz stability of Bayesian inverse problems. *Inverse Probl.*, 36(5):055015, 31, 2020. doi:10.1088/1361-6420/ab6f43.

A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numer.*, 19:451–559, 2010. doi:10.1017/S0962492910000061.

V. N. Sudakov. Linear sets with quasi-invariant measure. *Dokl. Akad. Nauk SSSR*, 127:524–525, 1959.