Edge-preserving Bayesian Inversion

Sergios Agapiou

Department of Mathematics and Statistics, University of Cyprus

Applied Inverse Problems 2017
29 May - 2 June 2017, Hangzhou

University of Cyprus

Outline

Outline

Inverse Problem

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

- $u \in X$ unknown function, X separable Banach space
- $y \in \mathbb{R}^J$ finite-dim observation
- $\xi \sim N(0, \Sigma)$, $\Sigma \in \mathbb{R}^{J \times J}$ positive definite, observational noise
- $\mathcal{G} : X \to \mathbb{R}^J$ possibly nonlinear forward operator, locally Lipschitz

Bayesian Formulation

- Prior: $u \sim \mu_0$
- Likelihood: $y|u \sim N(\mathcal{G}(u), \Sigma)$
- Posterior: $u|y \sim \mu^y$

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

$$\Phi(u;y) = \frac{1}{2}\left|\Sigma^{-\frac{1}{2}}(y - \mathcal{G}(u))\right|^2.$$

M. Dashti and A. M. Stuart, *The Bayesian approach to inverse problems*, Handbook of Uncertainty Quantification, 2015.

- We study MAP estimates in this non-parametric setting $\rightarrow$ modes of μ^y.

Edge-preserving and Sparsity-promoting Priors

Blocky structure and **sparsity** in an appropriate expansion

- 1-Besov priors
  - M. Lassas, E. Saksman and S. Siltanen, *Discretization-invariant Bayesian inversion and Besov space priors*, 2009
  - M. Dashti, S. Harris and A. Stuart, *Besov priors for Bayesian inverse problems*, 2013
- Infinitely divisible and heavy tailed priors, e.g. Cauchy priors
  - T. Sullivan, *Well-posed Bayesian inverse problems and heavy-tailed stable Banach space priors*, 2016
  - B. Hosseini, *Well-posed Bayesian inverse problems with infinitely-divisible and heavy-tailed prior measures*, 2017

http://www.sergiosagapiou.com/

S. Agapiou, M. Burger, M. Dashti and T. Helin, *Sparsity-promoting and edge-preserving maximum a posteriori estimators in non-parametric Bayesian inverse problems*, arXiv:1705.03286

Build on

M. Dashti, K. Law, A. Stuart and J. Voss, *MAP estimators and their consistency in Bayesian nonparametric inverse problems*, Inverse Problems, 2013

MAP for Gaussian priors

T. Helin and M. Burger, *Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems*, Inverse Problems, 2015

wMAP theory using differential calculus of measures, does not cover 1-Besov priors, basis for tackling Cauchy

Ongoing work on Cauchy priors with M. Dashti, T. Helin, H-C Lie and T. Sullivan

Outline

Finite-dimensional Intuition

- Assume $X = \mathbb{R}^N$ and prior has Lebesgue density

$$\pi(u) \propto \exp(-W(u))$$

- Posterior Lebesgue density

$$\pi^y(u) \propto \exp(-I(u; y)),$$

  where

$$I(u; y) = \Phi(u; y) + W(u).$$

- MAP estimators maximize posterior density, i.e. minimize Tikhonov functional I

Modes in Infinite-dimensions

- In infinite dimensions no uniform measure. Modes of measure μ on function space X?

  - compute $\mu(B_\epsilon(u))$ for all $u \in X$
  - send $\epsilon \to 0$
  - $\hat{u}$ mode of μ if it maximizes the limiting small ball probabilities in a specific sense

- strong mode: max probability among all centres in X, *Dashti et al '13*
- weak mode: max probability among all shifts of the ball by elements in a dense subspace $E \subset X$, *Helin and Burger '15*

MAP and wMAP

Definition (Dashti et al '13)

Let $M^\epsilon = \sup_{u \in X} \mu(B_\epsilon(u))$. $\hat{u} \in X$ is a mode of μ, if

$$\lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(\hat{u}))}{M^\epsilon} = 1.$$

Definition (Helin and Burger '15)

Let E dense subspace of X. $\hat{u} \in X$ weak mode of μ if

$$\lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(\hat{u} - h))}{\mu(B_\epsilon(\hat{u}))} \leq 1, \quad \forall h \in E.$$

A MAP (resp. wMAP) estimate is a mode (resp. weak mode) of μ^y.

Remarks

- Weak mode allows flexibility of choosing E.
- Any strong mode is a weak mode for $E = X$.
- Weak mode interesting when small ball probabilities available only in some subspace of translations h, E. Typically E has measure zero.

Strategy: Onsager-Machlup Functional

- Suppose can find $I : F \to \mathbb{R}$ s.t.

$$\lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(z_2))}{\mu(B_\epsilon(z_1))} = \exp(I(z_1) - I(z_2)).$$

- F dense subspace of X.
- Fix $z_1 \in F$. A $z_2 \in F$ minimizing I is a potential mode of μ.
- Such I is called the (generalized) Onsager-Machlup functional of μ.

Strategy: Onsager-Machlup Functional

- Suppose can find $I : F \to \mathbb{R}$ s.t.

$$\lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(z_2))}{\mu(B_\epsilon(z_1))} = \exp(I(z_1) - I(z_2)).$$

- F dense subspace of X.
- Fix $z_1 \in F$. A $z_2 \in F$ minimizing I is a potential mode of μ.
- Such I is called the (generalized) Onsager-Machlup functional of μ.
- For $X = \mathbb{R}^N$, $I = \Phi + W$ is the O-M functional of μ^y by Lebesgue differentiation theorem.

AIM: verify that wMAP and MAP identified with minimizers of O-M functional in ∞-dim.

Strategy: crucial first step

- For μ measure, define $\mu_h(\cdot) = \mu(\cdot - h)$.
- For h such that $\mu_h \ll \mu$, denote

$$R_h^\mu(u) = \frac{d\mu_h}{d\mu}(u).$$

Lemma (Helin and Burger '15)

$$\lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(u-h))}{\mu(B_\epsilon(u))} = R_h^\mu(u),$$

for all h such that R_h^μ is continuous for $u \in X$.

Proof.

$$\inf_{v \in B_\epsilon(u)} R_h^\mu(v) \leq \frac{\mu_h(B_\epsilon(u))}{\mu(B_\epsilon(u))} = \frac{\int_{B_\epsilon(u)} R_h^\mu(z)\mu(dz)}{\mu(B_\epsilon(u))} \leq \sup_{v \in B_\epsilon(u)} R_h^\mu(v),$$

for all $\epsilon > 0$ and $u \in X$. Take $\epsilon \to 0$ and use ctty. □

Strategy

- wMAP: choose E sufficiently regular s.t. $R_h^{\mu^y}$ continuous in $u \in X$, for all $h \in E$.
  - Straightforward to determine O-M functional I of μ^y, defined on F containing E.
  - Straightforward to study equivalence of wMAP and minimizers of I.
- MAP: considerably harder due to smallness of $F \subset X$ wrt prior hence posterior too.
- Suffices to establish analogous results for the prior μ_0 and use $\mu^y \propto e^{-\Phi}\mu_0$.

Strategy: establishing continuity of R_h^μ

- In HB15 ctty of R_h^μ established via differential calculus of measures assuming ctty of log-derivative β_h^μ for h in a sufficiently smooth space E.
  - Works for eg p-Besov priors with $1 < p \leq 2$ and Cauchy priors but not for 1-Besov prior whose log-derivative is inherently discontinuous.
- For product measures may be possible to get an explicit expression for R_h^μ via Kakutani-Hellinger theory and study its ctty analytically (works for 1-Besov priors)

Outline

Periodic Besov Spaces

- Let $\{\psi_\ell\}_{\ell=1}^{\infty}$ orthonormal wavelet basis for $L^2(\mathbb{T})$. Define $f : \mathbb{T} \to \mathbb{R}$ by

$$f(x) = \sum_{\ell=1}^{\infty} c_\ell \psi_\ell(x).$$

- $f \in B_p^s(\mathbb{T})$ iff

$$\|f\|_{B_p^s(\mathbb{T})} = \left(\sum_{\ell=1}^{\infty} \ell^{p(s+\frac{1}{2})-1} |c_\ell|^p \right)^{\frac{1}{p}} < \infty.$$

- p integrability, s smoothness parameter.
- For $p = 2$, Sobolev spaces of functions with s square integrable derivatives

$$\|f\|_{B_2^s(\mathbb{T})} = \left(\sum_{\ell=1}^{\infty} \ell^{2s} |c_\ell|^2 \right)^{\frac{1}{2}}.$$

- For $p = 1$

$$\|f\|_{B_1^s(\mathbb{T})} = \sum_{\ell=1}^{\infty} \ell^{s-\frac{1}{2}} |c_\ell|.$$

1-Besov Priors

Definition (Lassas et al '09)

Let $X_\ell \overset{iid}{\sim} \frac{1}{2}\exp(-|x|)$ and $\alpha_\ell = \ell^{s-\frac{1}{2}}$. The random function

$$U(x) = \sum_{\ell=1}^{\infty} \alpha_\ell^{-1} X_\ell \psi_\ell(x), \quad x \in \mathbb{T},$$

is said to be distributed according to a B_1^s-Besov prior, λ.

- Formally, U has density $\pi(u) \propto \exp(-\|u\|_{B_1^s})$ since $\alpha_\ell^{-1} X_\ell \sim \frac{\alpha_\ell}{2}\exp(-\alpha_\ell|x|)$.
- $\|U\|_{B_1^t} < \infty$, almost surely for all $t < s - 1$ (LSS09).
- $X = B_1^t$.

1-Besov Priors, R_h^λ

Lemma (A., Burger, Dashti and Helin '17)

- For all $h \in B_2^{s-\frac{1}{2}}$ we have $\lambda_h \ll \lambda$ and

$$\frac{d\lambda_h}{d\lambda}(u) = \lim_{N\to\infty} \exp \sum_{\ell=1}^{N} (-\alpha_\ell |h_\ell - u_\ell| + \alpha_\ell |u_\ell|).$$

- For $h \in B_1^r$, $r > s$, the limit on rhs is continuous in $u \in X = B_1^t$, $t < s - 1$.

Proof.

- Exploit product structure, use Hellinger integral to check equivalence of λ_h to λ. Kakutani theorem gives that

$$\frac{d\lambda_h}{d\lambda}(u) = \lim_{N\to\infty} \prod_{\ell=1}^{N} \frac{d\lambda_{h,\ell}}{d\lambda_\ell}(u) = \lim_{N\to\infty} \prod_{\ell=1}^{N} \frac{e^{-\alpha_\ell |h_\ell - u_\ell|}}{e^{-\alpha_\ell |u_\ell|}}.$$

- Technical explicit proof, showing that $|R_h^\mu(u) - R_h^\mu(v)| \to 0$ as $\|u - v\|_{B_1^t} \to 0$ by examining all combinations of signs.

□

1-Besov priors, Small Ball Probabilities

Corollary (A., Burger, Dashti and Helin '17)

$$\lim_{\epsilon \to 0} \frac{\lambda(B_\epsilon(u-h))}{\lambda(B_\epsilon(u))} = \exp \sum_{\ell=1}^{\infty} (-\alpha_\ell |h_\ell - u_\ell| + \alpha_\ell |u_\ell|),$$

for $h \in B_1^r$, $r > s$.

- Choose $E = B_1^r$, $r > s$ in wMAP definition (E dense in X).
- Sum makes sense for $u \in B_1^t$ provided $h \in B_1^r$, $r > s$. Can be split only for $u \in B_1^s$.
- To get Onsager-Machlup functional of λ need a bit more work.

1-Besov priors, Onsager-Machlup Functional of λ

Theorem (A., Burger, Dashti and Helin '17)

For $h \in B_1^s$,

$$\lim_{\epsilon \to 0} \frac{\lambda_h(B_\epsilon(0))}{\lambda(B_\epsilon(0))} = e^{-\|h\|_{B_1^s}}.$$

Proof.

For $h \in B_1^r$, $r > s$ it is immediate from last corollary. Take $h^j \subset B_1^{s+1}$ s.t. $h^j \to h$ in B_1^s. □

- For $z_1, z_2 \in B_1^s$

$$\lim_{\epsilon \to 0} \frac{\lambda(B_\epsilon(z_2))}{\lambda(B_\epsilon(z_1))} = \lim_{\epsilon \to 0} \frac{\lambda(B_\epsilon(z_2))}{\lambda(B_\epsilon(0))} \frac{\lambda(B_\epsilon(0))}{\lambda(B_\epsilon(z_1))} = \exp(\|z_1\|_{B_1^s} - \|z_2\|_{B_1^s}).$$

- The Onsager-Machlup functional of λ is $I : B_1^s \to [0, \infty)$, $I(z) = \|z\|_{B_1^s}$.

1-Besov Priors, MAP and wMAP

- Recall $\mu^y(du) \propto \exp(-\Phi(u;y))\lambda(du)$.
- Let

$$I(u;y) = \Phi(u;y) + \|u\|_{B_1^s}, \quad I : B_1^s \to [0,\infty).$$

- Finite-dim intuition suggests that minimizers of I are maximizers of posterior since formally

$$\text{"}\mu^y(du) \propto \exp(-\Phi(u;y) - \|u\|_{B_1^s})du\text{"}.$$

  We make this rigorous.

- Calculus of variations shows that I has a minimizer $\hat{u} \in B_1^s$ (ABDH17)

1-Besov priors, Onsager-Machlup Functional of μ^y

Proposition (A., Burger, Dashti and Helin '17)

I is the Onsager-Machlup functional for μ^y:

$$\lim_{\epsilon \to 0} \frac{\mu^y(B_\epsilon(z_2))}{\mu^y(B_\epsilon(z_1))} = \exp(I(z_1; y) - I(z_2; y)),$$

for all $z_1, z_2 \in B_1^s$.

1-Besov Priors, MAP and wMAP

Theorem (A., Burger, Dashti and Helin '17)

Both wMAP and MAP estimates of the posterior μ^y are identified with the minimizers of I.

Proof.

- wMAP:
  - $u_{min} \in B_1^s$ minimizer of I. For any $h \in B_1^r, r > s$ let $z_1 = u_{min} \in B_1^s$, $z_2 = u_{min} - h \in B_1^s$

$$\lim_{\epsilon \to 0} \frac{\mu^y(B_\epsilon(u_{min} - h))}{\mu^y(B_\epsilon(u_{min}))} \overset{prop}{=} \exp(I(u_{min}) - I(u_{min} - h)) \leq 1.$$

    Hence u_{min} is wMAP for $E = B_1^r, r > s$.
  - $\hat{u}$ wMAP for $E = B_1^r$. Exclude $\hat{u} \in B_1^t \setminus B_1^s$ by showing contradiction to ctty of Φ. Thus $\hat{u} \in B_1^s$ and by proposition

$$I(\hat{u}) - I(\hat{u} - h) \leq 0, \forall h \in B_1^r, r > s.$$

    By ctty of I in B_1^s and density of B_1^r in B_1^s, we get that $\hat{u}$ minimizer of I.

□

1-Besov Priors, MAP and wMAP

Theorem (A., Burger, Dashti and Helin '17)

Both wMAP and MAP estimates of the posterior μ^y are identified with the minimizers of I.

Proof.

- wMAP:
  - $u_{min} \in B_1^s$ minimizer of I. For any $h \in B_1^r, r > s$ let $z_1 = u_{min} \in B_1^s$, $z_2 = u_{min} - h \in B_1^s$

    $$\lim_{\epsilon \to 0} \frac{\mu^y(B_\epsilon(u_{min} - h))}{\mu^y(B_\epsilon(u_{min}))} \overset{prop}{=} \exp(I(u_{min}) - I(u_{min} - h)) \leq 1.$$

    Hence u_{min} is wMAP for $E = B_1^r, r > s$.
  - $\hat{u}$ wMAP for $E = B_1^r$. Exclude $\hat{u} \in B_1^t \setminus B_1^s$ by showing contradiction to ctty of Φ. Thus $\hat{u} \in B_1^s$ and by proposition

    $$I(\hat{u}) - I(\hat{u} - h) \leq 0, \forall h \in B_1^r, r > s.$$

    By ctty of I in B_1^s and density of B_1^r in B_1^s, we get that $\hat{u}$ minimizer of I.
- MAP: Very technical, requires many new small ball probability estimates for centres in B_1^t.

□

Outline

Conclusion and Future Work

- Wealth of new function-space priors, giving rise to many interesting questions.
- Discussed MAP estimates in this context, have a complete picture for Gaussian and 1-Besov priors, developing the picture for other priors.
- Other interesting questions:
  - When do MAP and wMAP coincide?
  - Local MAP and their theory
  - Consistency of MAP estimates in infinitely informative data limit (see DLSV13 and ABDH17)
  - Posterior contraction rates for the new priors (work in progress with M. Dashti and T. Helin for Besov priors)

http://www.sergiosagapiou.com/

- S. Agapiou, M. Burger, M. Dashti and T. Helin, *Sparsity-promoting and edge-preserving maximum a posteriori estimators in non-parametric Bayesian inverse problems*, arXiv:1705.03286
- M. Dashti, K. Law, A. Stuart and J. Voss, *MAP estimators and their consistency in Bayesian nonparametric inverse problems*, Inverse Problems, 2013
- T. Helin and M. Burger, *Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems*, Inverse Problems, 2015
- M. Dunlop and A. M. Stuart, *MAP estimators for piecewise continuous inversion*, Inverse Problems, 2016.
- M. Dashti and A. M. Stuart, *The Bayesian approach to inverse problems*, Handbook of Uncertainty Quantification, 2015.