Edge-preserving Bayesian Inversion

Sergios Agapiou

Department of Mathematics and Statistics, University of Cyprus

Bayesian and Nonlinear Inverse Problems
28 August - 1 September 2017, Leiden

University of Cyprus

Outline

1. Problem setup
2. 1-Besov priors
3. MAP and wMAP estimators
4. Conclusion

Outline

Inverse Problem

forward operator

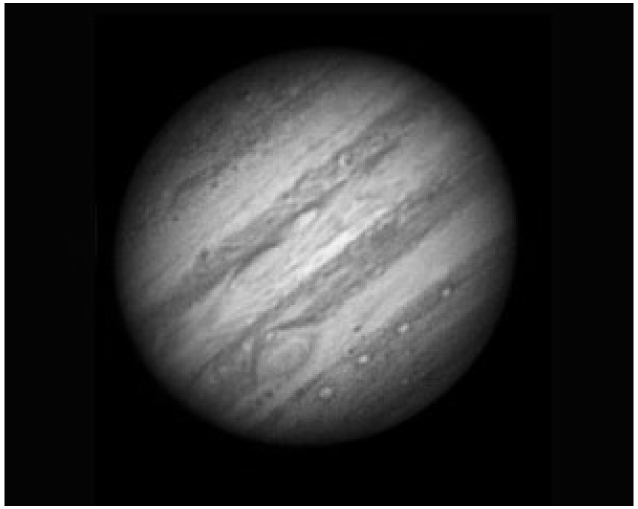
unknown ∞-dim

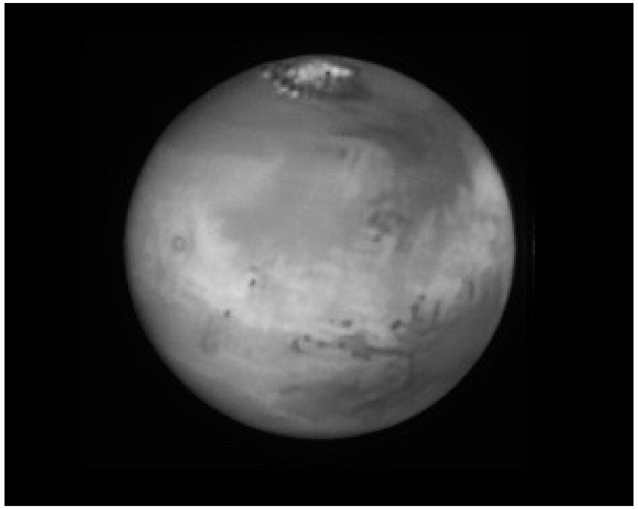
$$y = \mathcal{G}(u) + \xi$$

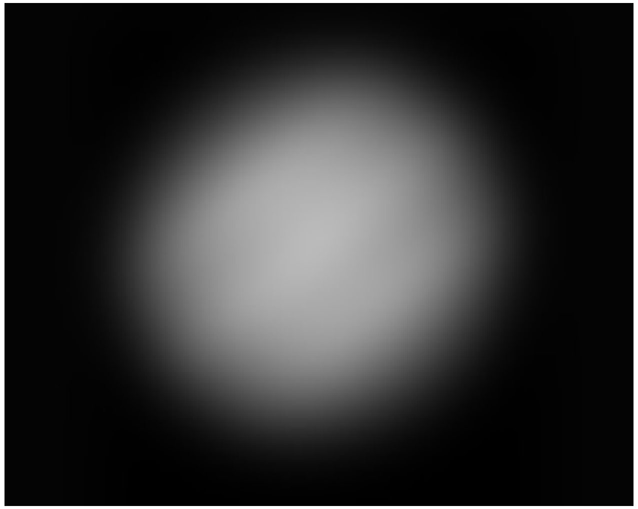
observation finite or ∞-dim

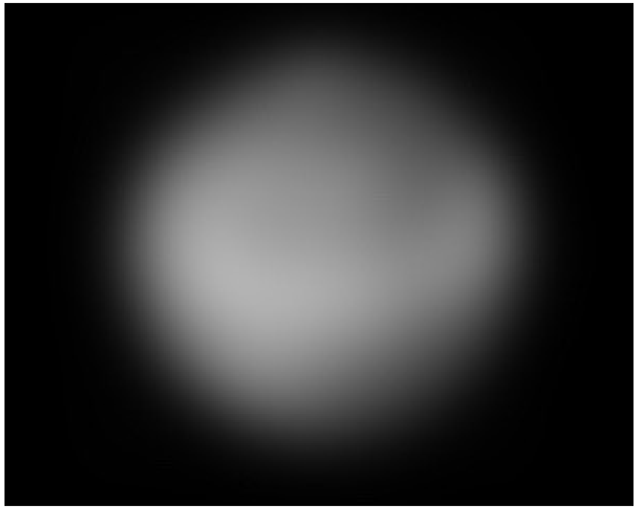
noise $N(0, \Sigma)$

Example - Image Deblurring

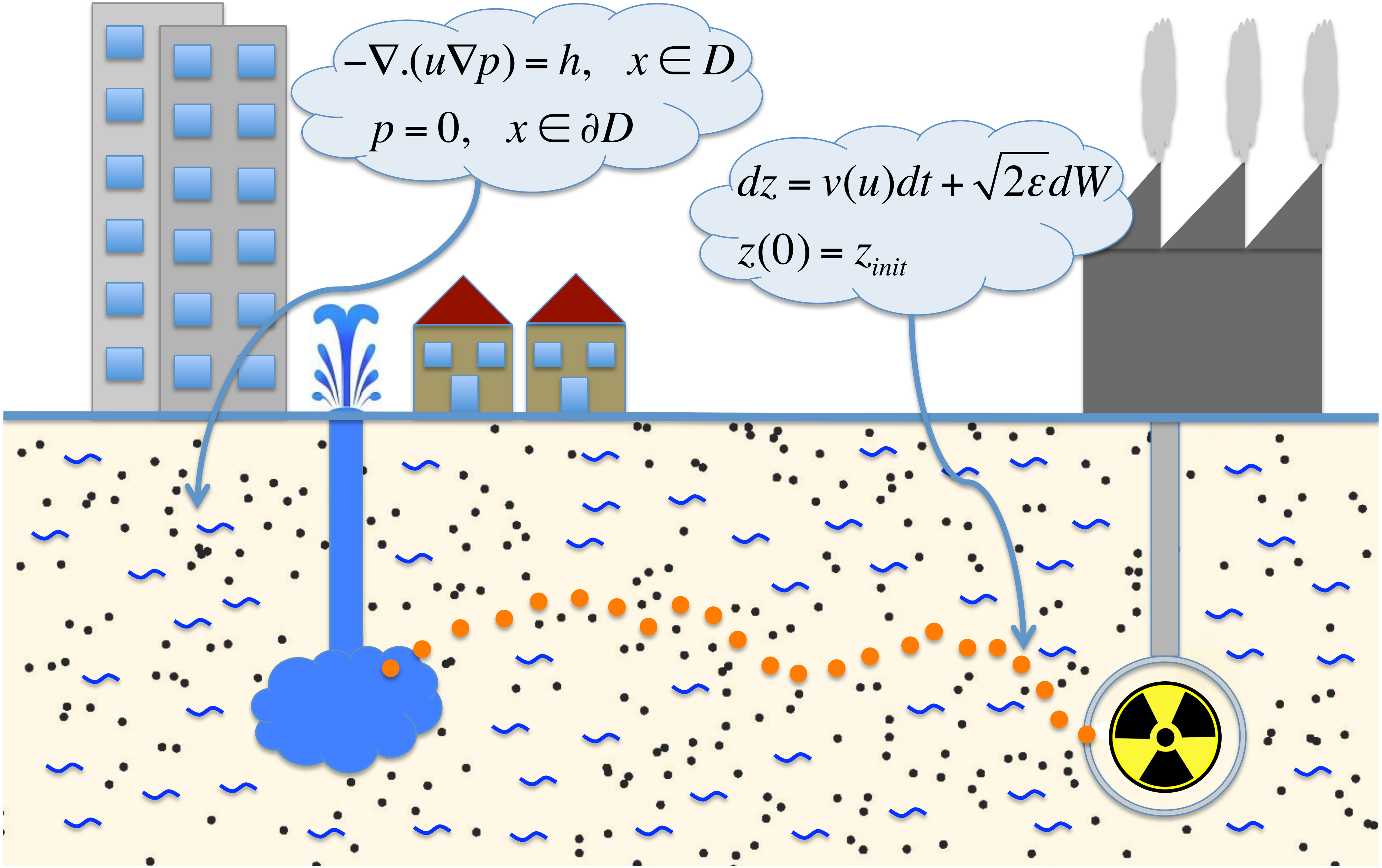








Example - Darcy Flow, Contamination Scenario



Classical Tikhonov-type Regularization

$$\hat{u} = \min I(u; y)$$

$$I(u; y) = \underbrace{\Phi(u; y)}_{\text{fidelity term}} + \underbrace{W(u)}_{\text{penalty term}}$$

Bayesian IP's for Functions, finite-dim observation

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 UQ, 2015.

Edge-preserving and Sparsity-promoting Priors

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

- Total Variation prior

$$"\mu_0(du) \propto \exp\left(-\int |Du|\right) du"$$

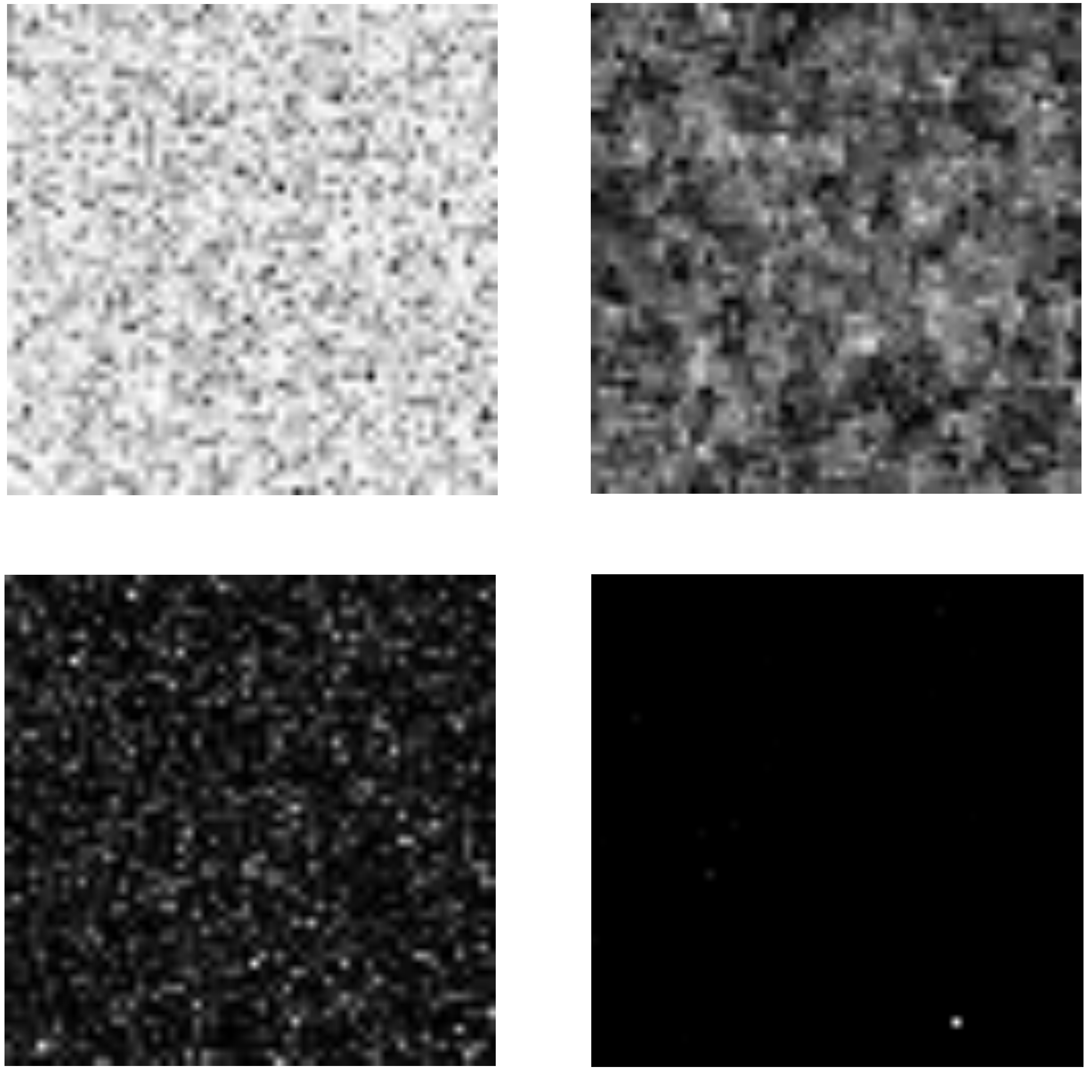
  - For $u = \mathbb{1}_A$, $\int |Du| = length(\partial A)$
  - Not discretization invariant

M. Lassas and S. Siltanen, *Can one use total variation prior for edge-preserving Bayesian inversion*, 2004

Edge-preserving and Sparsity-promoting Priors

- 1-Besov priors, Laplace-type, mimic TV.
  - 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

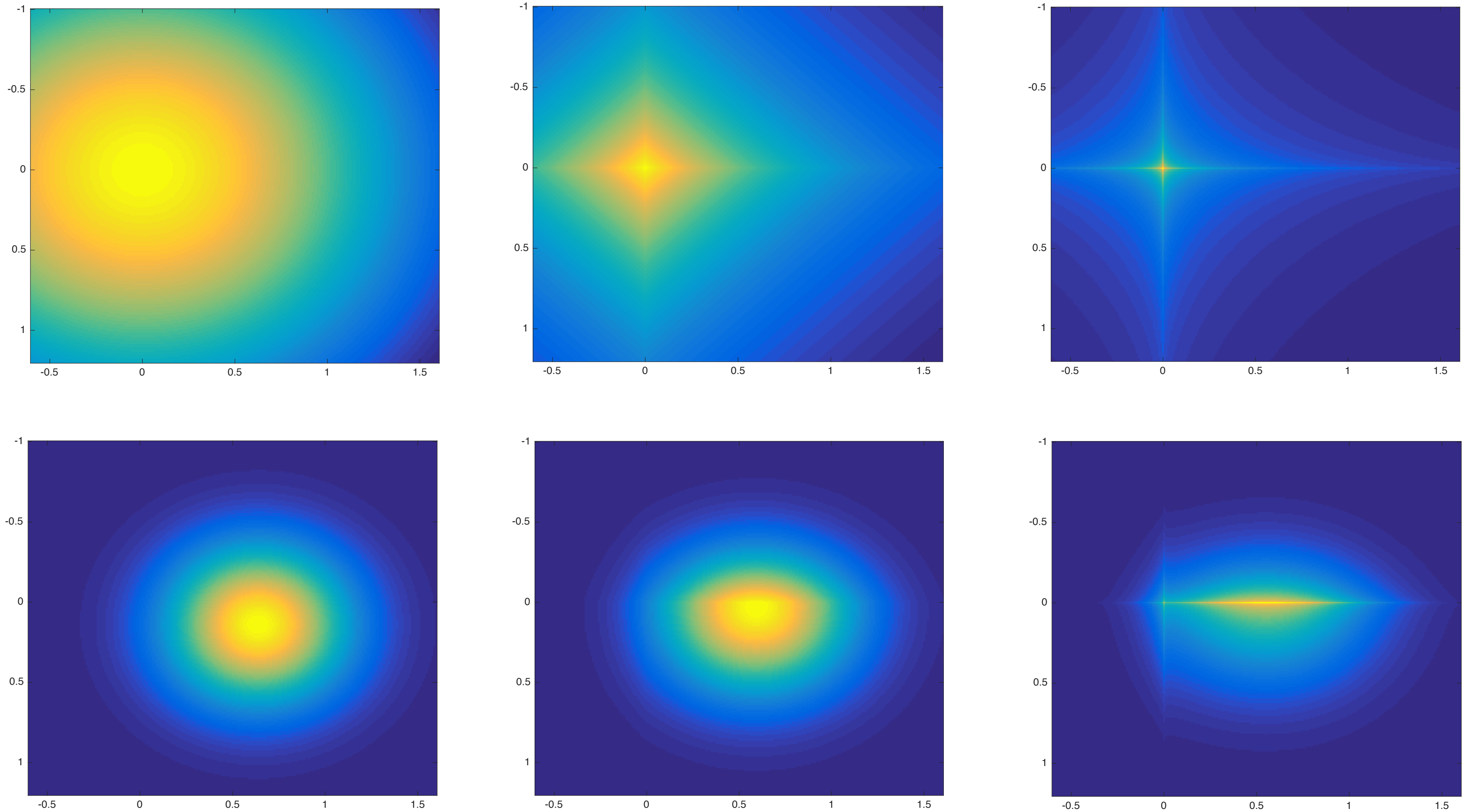
Priors for Blocky Structure and Sparsity



Gaussian, Total Variation, ℓ_1 and Cauchy draws

J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, 2005

$y = u + \xi$ in 2d



Gaussian ℓ_1 Cauchy

Outline

1. Problem setup
2. 1-Besov priors
3. MAP and wMAP estimators
4. Conclusion

Periodic Besov Spaces

- $\{\psi_\ell\}_{\ell=1}^{\infty}$ orthonormal wavelet basis for $L^2(\mathbb{T})$

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

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

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

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

- $p = 1$

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

1-Besov Priors

Definition (Lassas et al '09)

$X_\ell \overset{iid}{\sim} \frac{1}{2}\exp(-|x|)$ 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.

$$\lambda(B_1^t) = \begin{cases} 1, & \textit{if } t < s-1 \\ 0, & \textit{otherwise}. \end{cases}$$

- "$\pi_U(u) \propto \exp(-\|u\|_{B_1^s})$" since $\alpha_\ell^{-1} X_\ell \sim \frac{\alpha_\ell}{2}\exp(-\alpha_\ell|x|)$.

Study MAP Estimators

- Use 1-Besov priors in BIP context
- Study maximum a posteriori (MAP) estimators understood as modes of posterior μ^y

Outline

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 Cauchy

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 ∞-dim 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 maximizes limiting small ball probabilities in specific sense

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

AIM: associate abstract definitions to appropriate optimization problem.

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.

AIM: associate abstract definitions to appropriate optimization problem.

- B_1^s-Besov prior: show in ∞-dim that MAP/wMAP coincide with minimizers of

$$I(u; y) = \Phi(u; y) + \|u\|_{B_1^s}.$$

Strategy: Onsager-Machlup Functional

- Suppose can find $J : F \to [0, \infty)$ s.t.

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

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

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. $\square$

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

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

$I(u;y) = \Phi(u;y) + \|u\|_{B_1^s}$ is the Onsager-Machlup functional for μ^y, when $\mu_0 = \lambda$.

- Use Kakutani-Hellinger theory to get R_h^λ
- Check ctty of R_h^λ by brute force
- Density of Besov spaces arguments and $d\mu^y \propto e^{-\Phi} d\lambda$ give OM functional

Kakutani-Hellinger Theory

- For μ, ν measures both absolutely continuous wrt ζ, define Hellinger integral

$$H(\mu,\nu) = \int \sqrt{\frac{d\mu}{d\zeta}\frac{d\nu}{d\zeta}} d\zeta, \qquad H(\mu,\nu) \in [0,1].$$

- If $H(\mu,\nu) = 0$ then μ, ν singular.
- If $H(\mu,\nu) > 0$ then μ, ν not necessarily equivalent.

Kakutani-Hellinger Theory

If $\mu = \otimes_{\ell=1}^{\infty} \mu_\ell,\ \nu = \otimes_{\ell=1}^{\infty} \nu_\ell$, then

$$H(\mu, \nu) = \prod_{\ell=1}^{\infty} H(\mu_\ell, \nu_\ell).$$

Theorem (Kakutani)

Let μ, ν product measures, where μ_ℓ, ν_ℓ equivalent for all $\ell \in \mathbb{N}$. Then μ and ν equivalent iff $H(\mu, \nu) > 0$, and if equivalent

$$\frac{d\mu}{d\nu}(u) = \lim_{N\to\infty} \prod_{\ell=1}^{N} \frac{d\mu_\ell}{d\nu_\ell}(u_\ell), \quad \text{in } L^1(\mathbb{R}^\infty, \mu).$$

G. Da Prato, *An Introduction to Infinite-Dimensional Analysis*, Springer, 2006.

1-Besov Priors, R_h^λ

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

- We have $\lambda_h \sim \lambda$ if and only if $h \in B_2^{s-\frac{1}{2}}$.
- For $h \in B_2^{s-\frac{1}{2}}$

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

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

1-Besov Priors, R_h^λ

Proof.

- By Kakutani theorem suffices to compute $H(\lambda_h, \lambda)$ and find necessary and sufficient conditions on h ensuring its positivity.
- Kakutani theorem also gives that

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

- For ctty, 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.

$\square$

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

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

- With a little bit more work, can show that
  - OM functional of λ is $\|u\|_{B_1^s}$
  - OM functional of μ^y is $\Phi(u; y) + \|u\|_{B_1^s}$

1-Besov priors, Characterization of MAP and wMAP

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

Both wMAP and MAP estimates of the posterior μ^y identified with minimizers of $I(u;y) = \Phi(u;y) + \|u\|_{B_1^s}$.

- wMAP straightforward once we have OM functional, due to flexibility of choosing E
- MAP considerably harder, $\lambda(B_1^s) = 0$

1-Besov Priors, Consistency of MAP estimates

- Consider frequentist setup
$$y_j = \mathcal{G}(u^\dagger) + \xi_j,$$
for fixed underlying $u^\dagger \in X$ and $\xi_j \overset{i.i.d.}{\sim} N(0, \Sigma)$.

- Sequence of posteriors
$$\frac{d\mu^{y_1,\dots,y_n}}{d\lambda}(u) \propto \exp\Big(-\frac{1}{2}\sum_{j=1}^{n} |\Sigma^{-\frac{1}{2}}(y_j - \mathcal{G}(u))|^2\Big)$$

- Previous result shows that MAP (and wMAP) estimates coincide with minimizers of
$$I_n(u) = \frac{1}{2}\sum_{j=1}^{n} |\Sigma^{-\frac{1}{2}}(y_j - \mathcal{G}(u))|^2 + \|u\|_{B_1^s}.$$

1-Besov Priors, Consistency of MAP estimates

- Let $\{u_n\}$ be a sequence of MAP estimates corresponding to $\mu^{y_1,\dots,y_n}$.
- We investigate whether as $n \to \infty$, $\{u_n\}$ recovers $u^\dagger$ in some sense.
- Cannot expect to recover $u^\dagger$ fully, unless e.g. $\mathcal{G}$ is injective.

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

Assume $u^\dagger \in B_1^s$. Then there exists $u^\star \in B_1^s$ and a subsequence of $\{u_n\}$ such that $u_n \to u^\star$ in $B_1^{\tilde{s}}$ a.s., for any $\tilde{s} < s$. For any such $u^\star$ we have $\mathcal{G}(u^\star) = \mathcal{G}(u^\dagger)$.

Outline

Conclusion and Future Work

- Wealth of new function-space priors, many interesting questions arise.
- MAP in BIP context, complete picture for Gaussian and 1-Besov priors, developing the picture for other (Cauchy work in progress with Dashti, Helin)
- Other interesting questions:
  - When do MAP and wMAP coincide?
  - Local MAP and their theory
  - Posterior contraction rates for the new priors (work in progress with Dashti and 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.