

Bayesian inference and mathematical imaging. Part II: Markov chain Monte Carlo.

Dr. Marcelo Pereyra
<http://www.macs.hw.ac.uk/~mp71/>

Maxwell Institute for Mathematical Sciences, Heriot-Watt University

January 2019, IHP, Paris.



Outline

- 1 Bayesian inference in imaging inverse problems
- 2 Proximal Markov chain Monte Carlo
- 3 Uncertainty quantification in astronomical and medical imaging
- 4 Image model selection and model calibration
- 5 Conclusion

Imaging inverse problems

- We are interested in an unknown image $x \in \mathbb{R}^d$.
- We measure y , related to x by a statistical model $p(y|x)$.
- The recovery of x from y is ill-posed or ill-conditioned, **resulting in significant uncertainty about x .**
- For example, in many imaging problems

$$y = Ax + w,$$

for some operator A that is rank-deficient, and additive noise w .

The Bayesian framework

- We use priors to reduce uncertainty and deliver accurate results.
- Given the prior $p(x)$, the posterior distribution of x given y

$$p(x|y) = p(y|x)p(x)/p(y)$$

models our knowledge about x after observing y .

- In this talk we consider that $p(x|y)$ is log-concave; i.e.,

$$p(x|y) = \exp\{-\phi(x)\}/Z,$$

where $\phi(x)$ is a convex function and $Z = \int \exp\{-\phi(x)\}dx$.

Maximum-a-posteriori (MAP) estimation

The predominant Bayesian approach in imaging is MAP estimation

$$\begin{aligned}\hat{x}_{MAP} &= \operatorname{argmax}_{x \in \mathbb{R}^d} p(x|y), \\ &= \operatorname{argmin}_{x \in \mathbb{R}^d} \phi(x),\end{aligned}\tag{1}$$

computed efficiently, even in very high dimensions, by (proximal) convex optimisation (Chambolle and Pock, 2016).

Illustrative example: astronomical image reconstruction

Recover $x \in \mathbb{R}^d$ from low-dimensional degraded observation

$$y = M\mathcal{F}x + w,$$

where \mathcal{F} is the continuous Fourier transform, $M \in \mathbb{C}^{m \times d}$ is a measurement operator and w is Gaussian noise. We use the model

$$p(x|y) \propto \exp(-\|y - M\mathcal{F}x\|^2/2\sigma^2 - \theta\|\Psi x\|_1)\mathbf{1}_{\mathbb{R}_+^n}(x). \quad (2)$$

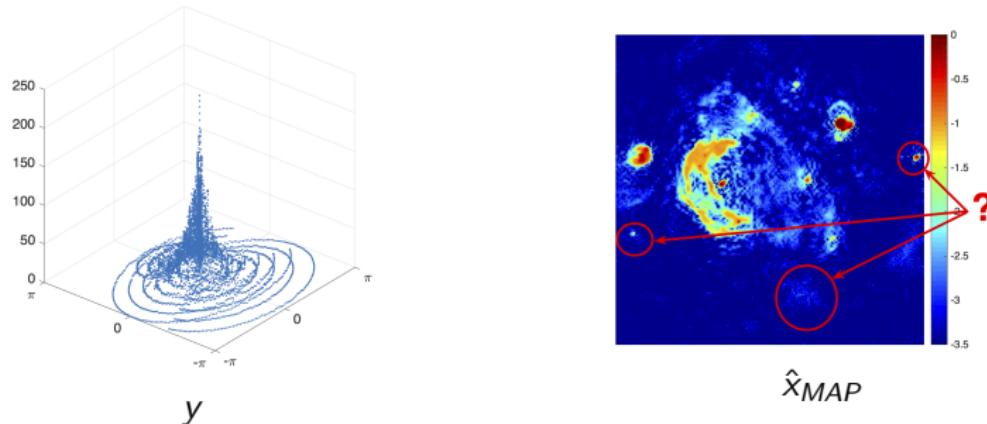


Figure : Radio-interferometric image reconstruction of the W28 supernova.

MAP estimation by proximal optimisation

To compute \hat{x}_{MAP} we use a proximal splitting algorithm. Let

$$f(x) = \|y - M\mathcal{F}x\|^2/2\sigma^2, \quad \text{and} \quad g(x) = \theta\|\Psi x\|_1 + -\log \mathbf{1}_{\mathbb{R}_+^n}(x),$$

where f and g are l.s.c. convex on \mathbb{R}^d , and f is L_f -Lipschitz differentiable.

For example, we could use a **proximal gradient** iteration

$$x^{m+1} = \text{prox}_g^{L_f^{-1}}\{x^m + L_f^{-1}\nabla f(x^m)\},$$

converges to \hat{x}_{MAP} at rate $O(1/m)$, with poss. acceleration to $O(1/m^2)$.

Definition For $\lambda > 0$, the λ -proximal operator of a convex l.s.c. function g is defined as (Moreau, 1962)

$$\text{prox}_g^\lambda(x) \triangleq \underset{u \in \mathbb{R}^N}{\operatorname{argmin}} g(u) + \frac{1}{2\lambda} \|u - x\|^2.$$

MAP estimation by proximal optimisation

The **alternating direction method of multipliers (ADMM)** algorithm

$$\begin{aligned}x^{m+1} &= \text{prox}_f^\lambda\{z^m - u^m\}, \\z^{m+1} &= \text{prox}_g^\lambda\{x^{m+1} + u^m\}, \\u^{m+1} &= u^m + x^{m+1} - z^{m+1},\end{aligned}$$

also converges to \hat{x}_{MAP} very quickly, and does not require f to be smooth.

However, MAP estimation has some limitations, e.g.,

- ① it provides little information about $p(x|y)$,
- ② it struggles with unknown/partially unknown models,

Outline

- 1 Bayesian inference in imaging inverse problems
- 2 Proximal Markov chain Monte Carlo
- 3 Uncertainty quantification in astronomical and medical imaging
- 4 Image model selection and model calibration
- 5 Conclusion

Inference by Markov chain Monte Carlo integration

Monte Carlo integration

Given a set of samples X_1, \dots, X_M distributed according to $p(x|y)$, we approximate posterior expectations and probabilities

$$\frac{1}{M} \sum_{m=1}^M h(X_m) \rightarrow E\{h(x)|y\}, \quad \text{as } M \rightarrow \infty$$

Markov chain Monte Carlo:

Construct a Markov kernel $X_{m+1}|X_m \sim K(\cdot|X_m)$ such that the Markov chain X_1, \dots, X_M has $p(x|y)$ as stationary distribution.

MCMC simulation in high-dimensional spaces is very challenging.

Unadjusted Langevin algorithm

Suppose for now that $p(x|y) \in \mathcal{C}^1$. Then, we can generate samples by mimicking a Langevin diffusion process that converges to $p(x|y)$ as $t \rightarrow \infty$,

$$\mathbf{X} : \quad d\mathbf{X}_t = \frac{1}{2} \nabla \log p(\mathbf{X}_t|y) dt + dW_t, \quad 0 \leq t \leq T, \quad \mathbf{X}(0) = x_0.$$

where W is the n -dimensional Brownian motion.

Because solving \mathbf{X}_t exactly is generally not possible, we use an [Euler Maruyama approximation](#) and obtain the “unadjusted Langevin algorithm”

$$\text{ULA} : \quad X_{m+1} = X_m + \delta \nabla \log p(X_m|y) + \sqrt{2\delta} Z_{m+1}, \quad Z_{m+1} \sim \mathcal{N}(0, \mathbb{I}_n)$$

ULA is remarkably efficient when $p(x|y)$ is sufficiently regular.

Non-smooth models

However, imaging models are often not smooth. Suppose that

$$p(x|y) \propto \exp \{-f(x) - g(x)\} \quad (3)$$

where $f(x)$ and $g(x)$ are l.s.c. convex functions from $\mathbb{R}^d \rightarrow (-\infty, +\infty]$, f is L_f -Lipschitz differentiable, and $g \notin \mathcal{C}^1$.

For example,

$$f(x) = \frac{1}{2\sigma^2} \|y - Ax\|_2^2, \quad g(x) = \alpha \|Bx\|_{\dagger} + \mathbf{1}_{\mathcal{S}}(x),$$

for some linear operators A , B , norm $\|\cdot\|_{\dagger}$, and convex set \mathcal{S} .

Unfortunately, such non-models are beyond the scope of ULA.

Idea: Regularise $p(x|y)$ to enable efficiently Langevin sampling.

Approximation of $p(x|y)$

Moreau-Yoshida approximation of $p(x|y)$ (Pereyra, 2015):

Let $\lambda > 0$. We propose to approximate $p(x|y)$ with the density

$$p_\lambda(x|y) = \frac{\exp[-f(x) - g_\lambda(x)]}{\int_{\mathbb{R}^d} \exp[-f(x) - g_\lambda(x)] dx},$$

where g_λ is the Moreau-Yoshida envelope of g given by

$$g_\lambda(x) = \inf_{u \in \mathbb{R}^d} \{g(u) + (2\lambda)^{-1} \|u - x\|_2^2\},$$

and where λ controls the approximation error involved.

Moreau-Yoshida approximations

Key properties (Pereyra, 2015; Durmus et al., 2018):

① $\forall \lambda > 0$, p_λ defines a proper density of a probability measure on \mathbb{R}^d .

② Convexity and differentiability:

- p_λ is log-concave on \mathbb{R}^d .
- $p_\lambda \in \mathcal{C}^1$ even if p not differentiable, with

$$\nabla \log p_\lambda(x|y) = -\nabla f(x) + \{\text{prox}_g^\lambda(x) - x\}/\lambda,$$

and $\text{prox}_g^\lambda(x) = \operatorname{argmin}_{u \in \mathbb{R}^N} g(u) + \frac{1}{2\lambda} \|u - x\|^2$.

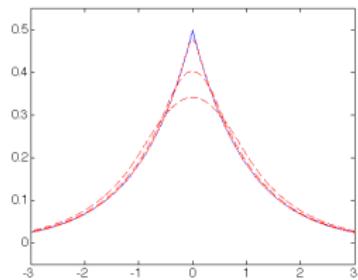
- $\nabla \log p_\lambda$ is Lipchitz continuous with constant $L \leq L_f + \lambda^{-1}$.

③ Approximation error between $p_\lambda(x|y)$ and $p(x|y)$:

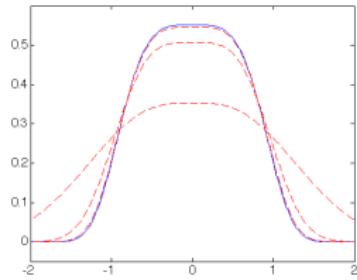
- $\lim_{\lambda \rightarrow 0} \|p_\lambda - p\|_{TV} = 0$.
- If g is L_g -Lipchitz, then $\|p_\lambda - p\|_{TV} \leq \lambda L_g^2$.

Illustration

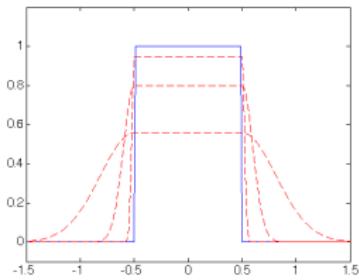
Examples of Moreau-Yoshida approximations:



$$p(x) \propto \exp(-|x|)$$



$$p(x) \propto \exp(-x^4)$$



$$p(x) \propto \mathbf{1}_{[-0.5,0.5]}(x)$$

Figure : True densities (solid blue) and approximations (dashed red).

Proximal ULA

We approximate \mathbf{X} with the “regularised” auxiliary Langevin diffusion

$$\mathbf{X}^\lambda : \quad d\mathbf{X}_t^\lambda = \frac{1}{2} \nabla \log p_\lambda(\mathbf{X}_t^\lambda | y) dt + dW_t, \quad 0 \leq t \leq T, \quad \mathbf{X}^\lambda(0) = x_0,$$

which targets $p_\lambda(x|y)$. Remark: we can make \mathbf{X}^λ arbitrarily close to \mathbf{X} .

Finally, an Euler Maruyama discretisation of \mathbf{X}^λ leads to the (Moreau-Yoshida regularised) proximal ULA

$$\text{MYULA} : \quad X_{m+1} = (1 - \frac{\delta}{\lambda}) X_m - \delta \nabla f\{X_m\} + \frac{\delta}{\lambda} \text{prox}_g^\lambda\{X_m\} + \sqrt{2\delta} Z_{m+1},$$

where we used that $\nabla g_\lambda(x) = \{x - \text{prox}_g^\lambda(x)\}/\lambda$.

Convergence results

Non-asymptotic estimation error bound

Theorem 2.1 (Durmus et al. (2018))

Let $\delta_\lambda^{\max} = (L_1 + 1/\lambda)^{-1}$. Assume that g is Lipschitz continuous. Then, there exist $\delta_\epsilon \in (0, \delta_\lambda^{\max}]$ and $M_\epsilon \in \mathbb{N}$ such that $\forall \delta < \delta_\epsilon$ and $\forall M \geq M_\epsilon$

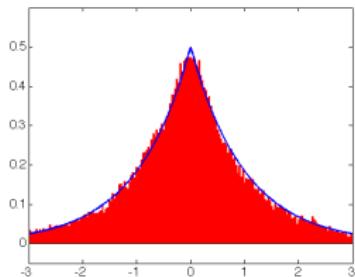
$$\|\delta_{x_0} Q_\delta^M - p\|_{TV} < \epsilon + \lambda L_g^2,$$

where Q_δ^M is the kernel assoc. with M iterations of MYULA with step δ .

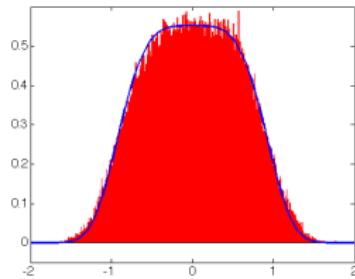
Note: δ_ϵ and M_ϵ are explicit and tractable. If $f + g$ is strongly convex outside some ball, then M_ϵ scales with order $\mathcal{O}(d \log(d))$. See Durmus et al. (2018) for other convergence results.

Illustration

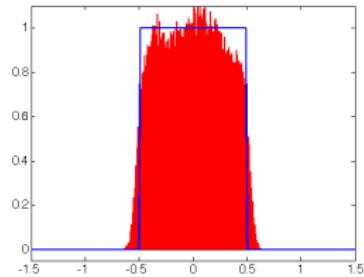
Illustrative examples:



$$p(x) \propto \exp(-|x|)$$



$$p(x) \propto \exp(-x^4)$$

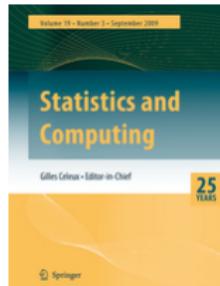


$$p(x) \propto \mathbf{1}_{[-0.5,0.5]}(x)$$

Figure : True densities (blue) and MC approximations (red histogram).

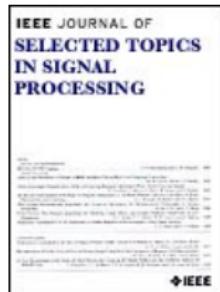
Modern Bayesian computation

Recent surveys on Bayesian computation...



25th anniversary special issue on Bayesian computation

P. Green, K. Latuszynski, M. Pereyra, C. P. Robert, "Bayesian computation: a perspective on the current state, and sampling backwards and forwards", *Statistics and Computing*, vol. 25, no. 4, pp 835-862, Jul. 2015.



Special issue on “Stochastic simulation and optimisation in signal processing”

M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tourneret, A. Hero, and S. McLaughlin, "A Survey of Stochastic Simulation and Optimization Methods in Signal Processing" *IEEE Sel. Topics in Signal Processing*, vol. 10, no. 2, pp 224 - 241, Mar. 2016.

Outline

- 1 Bayesian inference in imaging inverse problems
- 2 Proximal Markov chain Monte Carlo
- 3 Uncertainty quantification in astronomical and medical imaging
- 4 Image model selection and model calibration
- 5 Conclusion

Uncertainty quantification in radio-interferometric imaging

Where does the posterior probability mass of x lie?

- A set C_α is a posterior credible region of confidence level $(1 - \alpha)\%$ if

$$P[x \in C_\alpha | y] = 1 - \alpha.$$

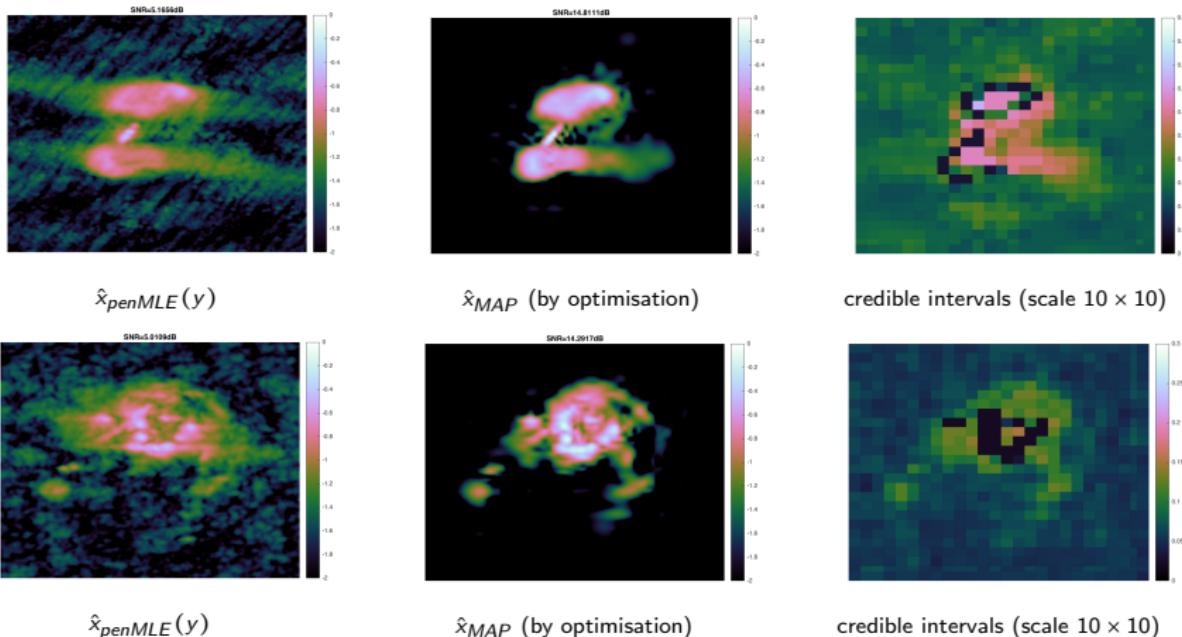
- The *highest posterior density* (HPD) region is decision-theoretically optimal (Robert, 2001)

$$C_\alpha^* = \{x : \phi(x) \leq \gamma_\alpha\}$$

with $\gamma_\alpha \in \mathbb{R}$ chosen such that $\int_{C_\alpha^*} p(x|y) dx = 1 - \alpha$ holds.

Visualising uncertainty in radio-interferometric imaging

Astro-imaging experiment with redundant wavelet frame (Cai et al., 2017).



3C2888 and M31 radio galaxies (size 256×256 pixels). Estimation error w.r.t.
MH implementation 3%.

Hypothesis testing

Bayesian hypothesis test for specific image structures (e.g., lesions)

H_0 : The structure of interest is ABSENT in the true image

H_1 : The structure of interest is PRESENT in the true image

The null hypothesis H_0 is rejected with significance α if

$$P(H_0|y) \leq \alpha.$$

Theorem (Repetti et al., 2018)

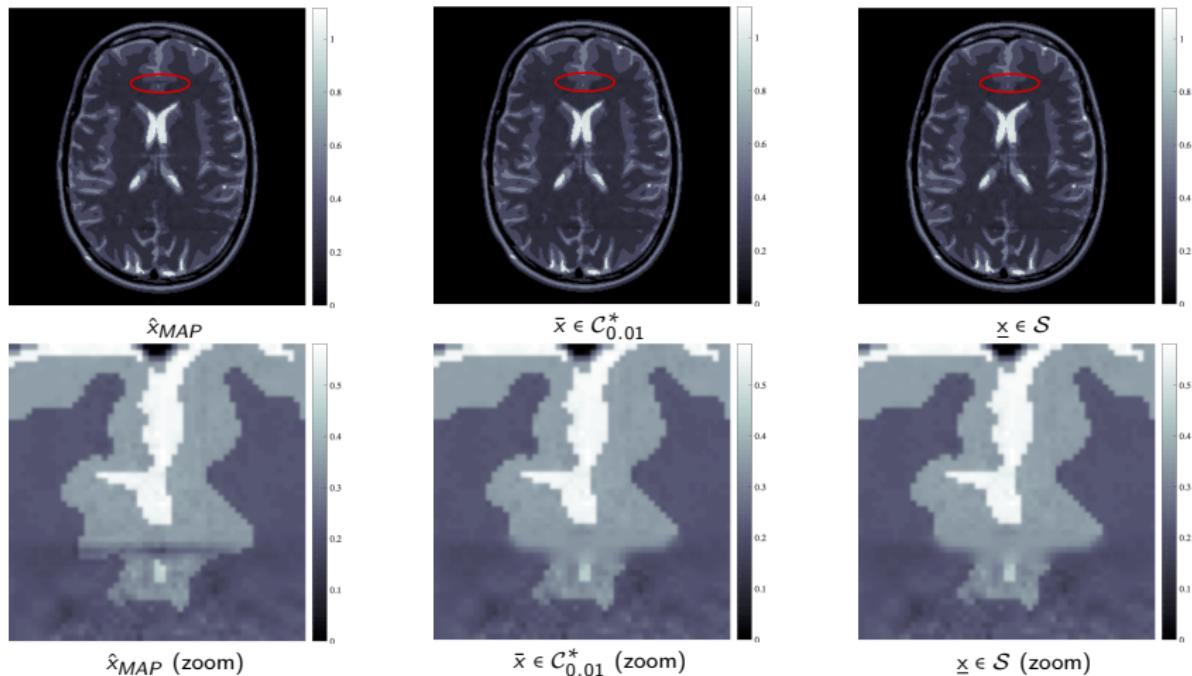
Let \mathcal{S} denote the region of \mathbb{R}^d associated with H_0 , containing all images *without the structure* of interest. Then

$$\mathcal{S} \cap \mathcal{C}_\alpha^* = \emptyset \implies P(H_0|y) \leq \alpha.$$

If in addition \mathcal{S} is convex, then checking $\mathcal{S} \cap \mathcal{C}_\alpha^* = \emptyset$ is a convex problem

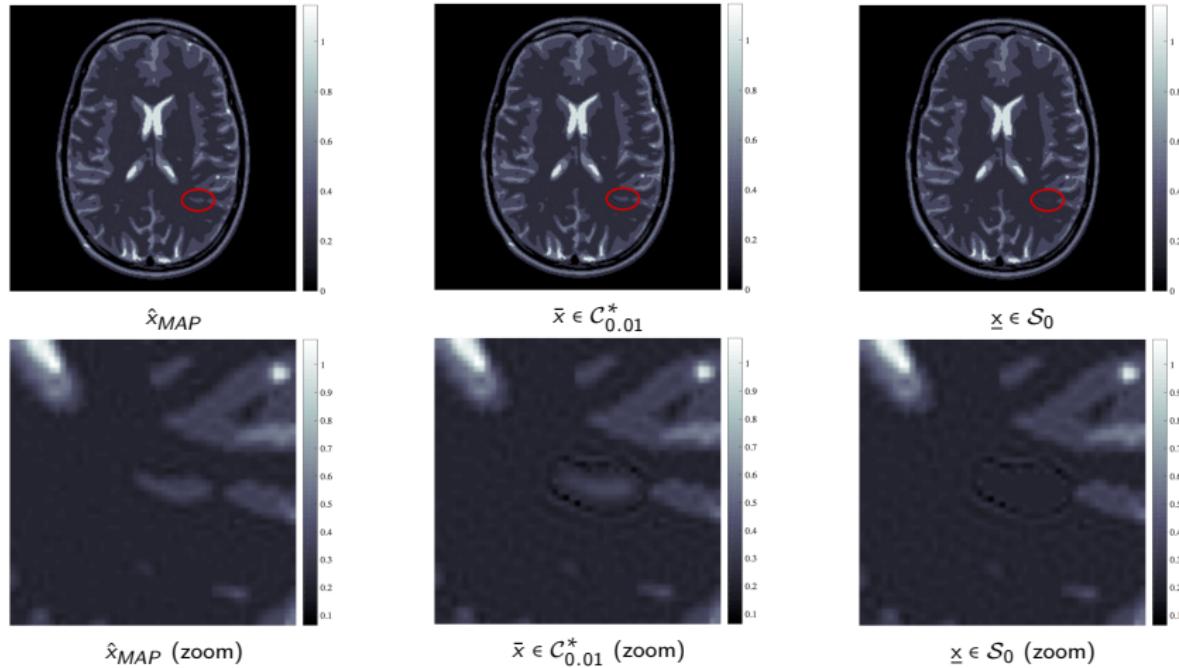
$$\min_{\bar{x}, \underline{x} \in \mathbb{R}^d} \|\bar{x} - \underline{x}\|_2^2 \quad \text{s.t.} \quad \bar{x} \in \mathcal{C}_\alpha^*, \quad \underline{x} \in \mathcal{S}.$$

Uncertainty quantification in MRI imaging



MRI experiment: test images $\bar{x} = \underline{x}$, hence we fail to reject H_0 and conclude that there is little evidence to support the observed structure.

Uncertainty quantification in MRI imaging



MRI experiment: test images $\bar{x} \neq \underline{x}$, hence we reject H_0 and conclude that there is significant evidence in favour of the observed structure.

Uncertainty quantification in radio-interferometric imaging

Quantification of minimum energy of different energy structures, at level $(1 - \alpha) = 0.99$, as the number of measurements $T = \dim(y)/2$ increases.

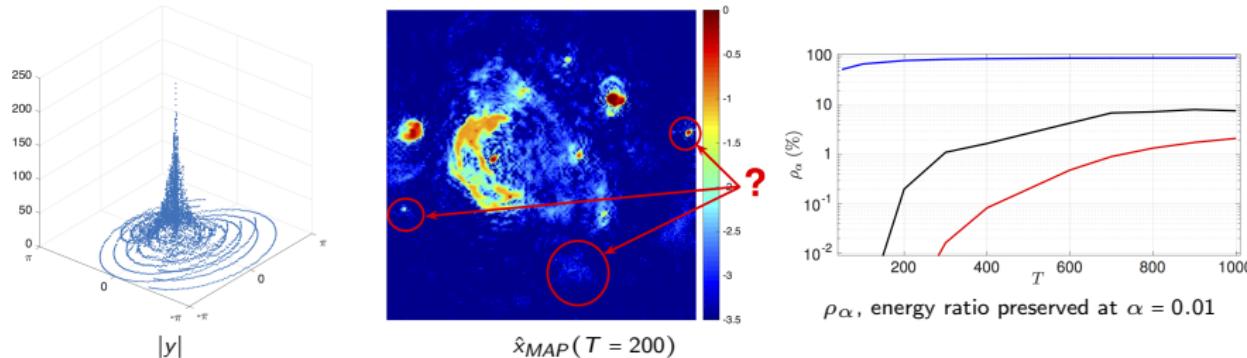


Figure : Analysis of 3 structures in the W28 supernova RI image.

Note: energy ratio calculated as

$$\rho_\alpha = \frac{\|\bar{x} - \underline{x}\|_2}{\|x_{MAP} - \tilde{x}_{MAP}\|_2}$$

where \bar{x}, \underline{x} are computed with $\alpha = 0.01$, and \tilde{x}_{MAP} is a modified version of x_{MAP} where the structure of interest has been carefully removed from the image.

Outline

- 1 Bayesian inference in imaging inverse problems
- 2 Proximal Markov chain Monte Carlo
- 3 Uncertainty quantification in astronomical and medical imaging
- 4 Image model selection and model calibration
- 5 Conclusion

Bayesian Model Selection

The Bayesian framework provides theory for comparing models objectively.

Given K alternative models $\{\mathcal{M}_j\}_{j=1}^K$ with posterior densities

$$\mathcal{M}_j : p_j(x|y) = p_j(y|x)p_j(x)/p_j(y),$$

we compute the (marginal) posterior probability of each model, i.e.,

$$p(\mathcal{M}_j|y) \propto p(y|\mathcal{M}_j)p(\mathcal{M}_j) \quad (4)$$

where $p(y|\mathcal{M}_j) \doteq p_j(y) = \int p_j(y|x)p_j(x)dx$ measures model-fit-to-data.

We then **select** for our inferences the “best” model, i.e.,

$$\mathcal{M}^* = \operatorname{argmax}_{j \in \{1, \dots, K\}} p(\mathcal{M}_j|y).$$

Experiment setup

We degrade the Boat image of size 256×256 pixels with a 5×5 uniform blur operator A^* and Gaussian noise $w \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_N)$ with $\sigma = 0.5$.

$$y = A^*x + w$$

We consider four alternative models to estimate x , given by

$$\mathcal{M}_j : p_j(x|y) \propto \exp \left[-(\|y - A_j x\|^2 / 2\sigma^2) - \beta_j \phi_j(x) \right] \quad (5)$$

with fixed hyper-parameters σ and β , and where:

- \mathcal{M}_1 : A_1 is the **correct** blur operator and $\phi_j(x) = TV(x)$.
- \mathcal{M}_2 : A_2 is a **mildly misspecified** blur operator and $\phi_j(x) = TV(x)$.
- \mathcal{M}_3 : A_3 is the **correct** blur operator and $\phi_j(x) = \|\Psi x\|_1$.
- \mathcal{M}_4 : A_4 is a **mildly misspecified** blur operator and $\phi_j(x) = \|\Psi x\|_1$.

where Ψ is a wavelet frame and $TV(x) = \|\nabla_d x\|_{1-2}$ is the total-variation pseudo-norm. The β_j are adjusted automatically (see model calibration).

Monte Carlo strategy

To perform model selection we use MYULA to approximate the posterior probabilities $p(\mathcal{M}_j|y)$ for $j = 1, 2, 3, 4$ by Monte Carlo integration.

For each model we generate $n = 10^5$ samples $\{X_k^j\}_{k=1}^n \sim p(x|y, \mathcal{M}_j)$ and use the truncated harmonic mean estimator

$$p(y|\mathcal{M}_j) \approx \left(\sum_{k=1}^n \frac{\mathbf{1}_{\mathcal{S}^*}(X_k^M)}{p(X_k^M, y|\mathcal{M}_j)} \right)^{-1} \text{vol}(\mathcal{S}^*), \quad j = \{1, 2, 3, 4\} \quad (6)$$

where \mathcal{S}^* is a union of highest posterior density sets of $p(x|y, \mathcal{M}_j)$, also estimated from $\{X_k^j\}_{k=1}^n$.

Computing time approx. 30 minutes per model.

Numerical results

We obtain that $p(\mathcal{M}_1|y) \approx 0.68$ and $p(\mathcal{M}_3|y) \approx 0.27$ with the **correct blur are the best models**, $p(\mathcal{M}_2|y) < 0.05$ and $p(\mathcal{M}_4|y) < 0.01$ perform poorly.

\mathcal{M}_1



y

\mathcal{M}_3



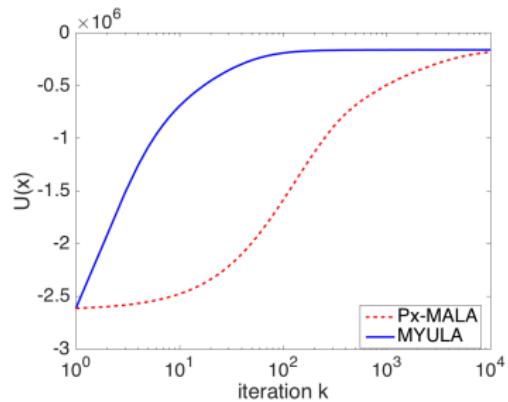
\hat{x}_{MAP} (PSNR 32.9dB)
 $p(\mathcal{M}_3|y) \approx 0.27$

\hat{x}_{MAP} (PSNR 34.1dB)
 $p(\mathcal{M}_1|y) \approx 0.68$

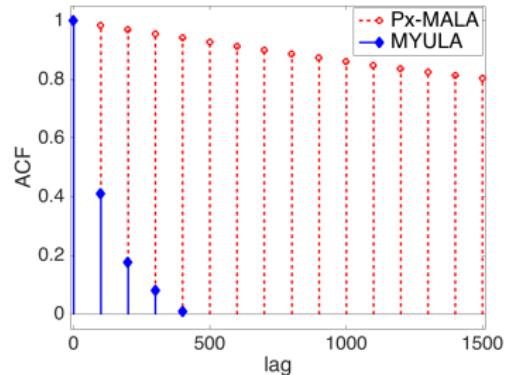
Figure : MAP estimation results for the Boat image deblurring experiment.
(Note: error w.r.t. "exact" probabilities from Px-MALA approx. 0.5%.)

Numerical results

MYULA and Px-MALA efficiency comparison:



(a)



(b)

Figure : (a) Convergence of the chains to the typical set of $x|y$ under model \mathcal{M}_1 ,
(b) chain autocorrelation function (ACF).)

Empirical Bayesian model calibration

For illustration, consider the class of Bayesian models

$$p(x|y, \theta) = \frac{p(y|x)p(x|\theta)}{p(y|\theta)},$$

parametrised by a **regularisation parameter** $\theta \in \Theta$. For example,

$$p(x|\theta) = \frac{1}{C(\theta)} \exp\{-\theta \varphi(x)\}, \quad p(y|x) \propto \exp\{-f_y(x)\},$$

with f_y and φ convex l.s.c. functions, and f_y L -Lipschitz differentiable.

We assume that $p(x|\theta)$ is proper, i.e.,

$$C(\theta) = \int_{\mathbb{R}^d} \exp\{-\theta \varphi(x)\} dx < \infty,$$

with $C(\theta)$ unknown and generally intractable.

Maximum-a-posteriori estimation

If θ is fixed, the posterior $p(x|y, \theta)$ is log-concave and

$$\hat{x}_{MAP} = \operatorname{argmin}_{x \in \mathbb{R}^d} f_y(x) + \theta \varphi(x)$$

is a convex optimisation problem that can be often solved efficiently.

For example, the proximal gradient algorithm

$$x^{m+1} = \operatorname{prox}_{\varphi}^{L^{-1}} \{x^m + L^{-1} \nabla f_y(x^m)\},$$

converges to \hat{x}_{MAP} as $m \rightarrow \infty$.

However, when θ is unknown this significantly complicates the problem.

Regularisation parameter MLE

We adopt an empirical Bayes approach and calibrate the model maximising the evidence or marginal likelihood, i.e.,

$$\begin{aligned}\hat{\theta} &= \operatorname{argmax}_{\theta \in \Theta} p(y|\theta), \\ &= \operatorname{argmax}_{\theta \in \Theta} \int_{\mathbb{R}^d} p(y, x|\theta) dx,\end{aligned}$$

which we solve efficiently by using a stochastic gradient algorithm driven by two proximal MCMC kernels (see Fernandez-Vidal and Pereyra (2018)).

Given $\hat{\theta}$, we then straightforwardly compute

$$\hat{x}_{MAP} = \operatorname{argmin}_{x \in \mathbb{R}^d} f_y(x) + \hat{\theta} \varphi(x). \quad (7)$$

Projected gradient algorithm

Assume that Θ is convex, and that $\hat{\theta}$ is the only root of $\nabla_{\theta} \log p(y|\theta)$ in Θ . Then $\hat{\theta}$ is also the unique solution of the fixed-point equation

$$\theta = P_{\Theta} [\theta + \delta \nabla_{\theta} \log p(y|\theta)] .$$

where P_{Θ} is the projection operator on Θ and $\delta > 0$.

If $\nabla \log p(y|\theta)$ was tractable, we could compute $\hat{\theta}$ iteratively by using

$$\theta^{(t+1)} = P_{\Theta} \left[\theta^{(t)} + \delta_t \nabla_{\theta} \log p(y|\theta^{(t)}) \right],$$

with sequence $\delta_t = \alpha t^{-\beta}$, $\alpha > 0$, $\beta \in [1/2, 1]$.

However, $\nabla \log p(y|\theta)$ is “doubly” intractable...

Stochastic projected gradient algorithm

To circumvent the intractability of $\nabla_{\theta} \log p(y|\theta)$ we use Fisher's identity

$$\begin{aligned}\nabla_{\theta} \log p(y|\theta) &= E_{x|y,\theta} \{ \nabla_{\theta} \log p(x,y|\theta) \}, \\ &= -E_{x|y,\theta} \{ \varphi + \nabla_{\theta} \log C(\theta) \},\end{aligned}$$

together with the identity

$$\nabla_{\theta} \log C(\theta) = -E_{x|\theta} \{ \varphi(x) \},$$

to obtain $\nabla_{\theta} \log p(y|\theta) = E_{x|\theta} \{ \varphi(x) \} - E_{x|y,\theta} \{ \varphi(x) \}$.

This leads to the equivalent fixed-point equation

$$\theta = P_{\Theta} (\theta + \delta E_{x|\theta} \{ \varphi(x) \} - \delta E_{x|y,\theta} \{ \varphi(x) \}), \quad (8)$$

which we solve by using a stochastic approximation algorithm.

Stochastic Approximation algorithm to compute $\hat{\theta}$

We use the following MCMC-driven stochastic gradient algorithm:

Initialisation $x^{(0)}, u^{(0)} \in \mathbb{R}^d$, $\theta^{(0)} \in \Theta$, $\delta_t = \delta_0 t^{-0.8}$.

for $t = 0$ to n

1. MCMC update $x^{(t+1)} \sim M_{x|y,\theta^{(t)}}(\cdot|x^{(t)})$ targeting $p(x|y, \theta^{(t)})$
2. MCMC update $u^{(t+1)} \sim K_{x|\theta^{(t)}}(\cdot|u^{(t)})$ targeting $p(x|\theta^{(t)})$
3. Stoch. grad. update

$$\theta^{(t+1)} = P_\Theta \left[\theta^{(t)} + \delta_t \varphi(u^{(t+1)}) - \delta_t \varphi(x^{(t+1)}) \right].$$

end for

Output The iterates $\theta^{(t)} \rightarrow \hat{\theta}$ as $n \rightarrow \infty$.

SAPG algorithm driven MCMC kernels

Initialisation $x^{(0)}, u^{(0)} \in \mathbb{R}^d$, $\theta^{(0)} \in \Theta$, $\delta_t = \delta_0 t^{-0.8}$, $\lambda = 1/L$, $\gamma = 1/4L$.

for $t = 0$ to n

1. Coupled Proximal MCMC updates: generate $z^{(t+1)} \sim \mathcal{N}(0, \mathbb{I}_d)$

$$\begin{aligned} x^{(t+1)} &= \left(1 - \frac{\gamma}{\lambda}\right) x^{(t)} - \gamma \nabla f_y \left(x^{(t)} \right) + \frac{\gamma}{\lambda} \text{prox}_{\varphi}^{\theta\lambda} \left(x^{(t)} \right) + \sqrt{2\gamma} z^{(t+1)}, \\ u^{(t+1)} &= \left(1 - \frac{\gamma}{\lambda}\right) u^{(t)} + \frac{\gamma}{\lambda} \text{prox}_{\varphi}^{\theta\lambda} \left(u^{(t)} \right) + \sqrt{2\gamma} z^{(t+1)}, \end{aligned}$$

2. Stochastic gradient update

$$\theta^{(t+1)} = P_{\Theta} \left[\theta^{(t)} + \delta_t \varphi(u^{(t+1)}) - \delta_t \varphi(x^{(t+1)}) \right].$$

end for

Output Averaged estimator $\bar{\theta} = n^{-1} \sum_{t=1}^n \theta^{(t+1)}$ converges approx. to $\hat{\theta}$.

Illustrative example - Image deblurring with ℓ_1 prior

We consider again the live-cell microscopy setup

$$p(x|y, \theta) \propto \exp(-\|y - Ax\|^2/2\sigma^2 - \theta\|x\|_1),$$

and compute $\hat{\theta} = \operatorname{argmax}_{\theta \in \mathbb{R}^+} p(y|\theta)$.

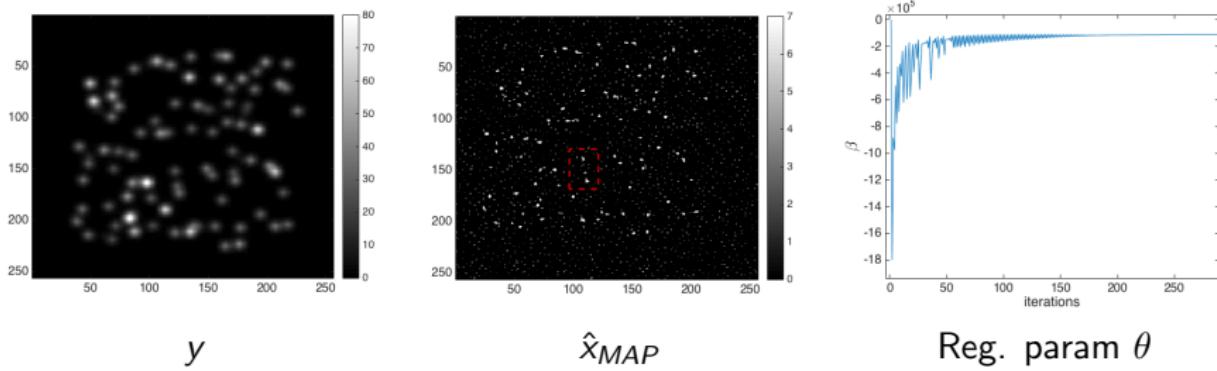


Figure : Molecules image deconvolution experiment, computing time 0.75 secs.

Illustrative example - Image deblurring with TV- ℓ_2 prior

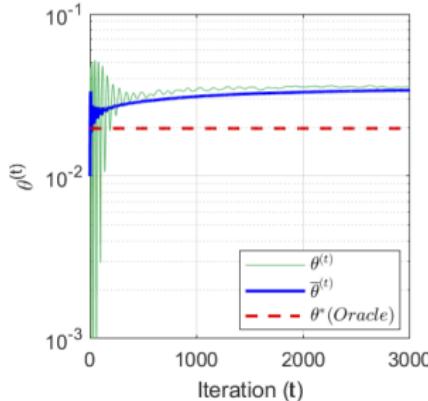
Similarly, for the Bayesian image deblurring model

$$p(x|y, \theta) \propto \exp\left(-\|y - Ax\|^2/2\sigma^2 - \alpha\|x\|_2 - \theta\|\nabla_d x\|_{1-2}\right),$$

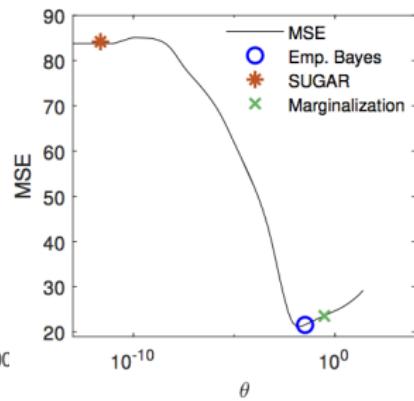
we compute $\hat{\theta} = \operatorname{argmax}_{\theta \in \mathbb{R}^+} p(y|\theta)$.



y



Reg. param θ



Estimation error for \hat{x}_{MAP}

Figure : Boat image deconvolution experiment.

Image deblurring with TV- ℓ_2 prior

Comparison with the (non-Bayesian) SUGAR method (Deledalle et al., 2014), and an oracle that knows the optimal value of θ . Average values over 6 test images of size 512×512 pixels.



(a) Original

(b) Degraded

(c) Emp. Bayes

(d) SUGAR

Method	SNR=20 dB				SNR=30 dB				SNR=40 dB			
	Avg.	MSE	Avg.	Time	Avg.	MSE	Avg.	Time	Avg.	MSE	Avg.	Time
θ^* (Oracle)	22.95	\pm 3.10	—	—	21.05	\pm 3.19	—	—	18.76	\pm 3.19	—	—
Empirical Bayes	23.24	\pm 3.23	43.01	—	21.16	\pm 3.24	41.50	—	18.90	\pm 3.39	42.85	—
SUGAR	24.14	\pm 3.19	15.74	—	23.96	\pm 3.26	20.87	—	23.94	\pm 3.27	20.59	—

Outline

- 1 Bayesian inference in imaging inverse problems
- 2 Proximal Markov chain Monte Carlo
- 3 Uncertainty quantification in astronomical and medical imaging
- 4 Image model selection and model calibration
- 5 Conclusion

Conclusion

- The challenges facing modern imaging sciences require a methodological paradigm shift to go beyond point estimation.
- The Bayesian framework can support this paradigm shift, but this requires significantly accelerating computation methods.
- We explored improving efficiency by integrating modern stochastic and variational approaches.

Thank you!

Bibliography:

- Cai, X., Pereyra, M., and McEwen, J. D. (2017). Uncertainty quantification for radio interferometric imaging II: MAP estimation. *ArXiv e-prints*.
- Chambolle, A. and Pock, T. (2016). An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319.
- Deledalle, C.-A., Vaiter, S., Fadili, J., and Peyré, G. (2014). Stein unbiased gradient estimator of the risk (sugar) for multiple parameter selection. *SIAM Journal on Imaging Sciences*, 7(4):2448–2487.
- Durmus, A., Moulines, E., and Pereyra, M. (2018). Efficient Bayesian computation by proximal Markov chain Monte Carlo: when Langevin meets Moreau. *SIAM J. Imaging Sci.*, 11(1):473–506.
- Fernandez-Vidal, A. and Pereyra, M. (2018). Maximum likelihood estimation of regularisation parameters. In *Proc. IEEE ICIP 2018*.
- Moreau, J.-J. (1962). Fonctions convexes duales et points proximaux dans un espace Hilbertien. *C. R. Acad. Sci. Paris Sér. A Math.*, 255:2897–2899.
- Pereyra, M. (2015). Proximal Markov chain Monte Carlo algorithms. *Statistics and Computing*. open access paper, <http://dx.doi.org/10.1007/s11222-015-9567-4>.
- Repetti, A., Pereyra, M., and Wiaux, Y. (2018). Scalable Bayesian uncertainty quantification in imaging inverse problems via convex optimisation. *ArXiv e-prints*.
- Robert, C. P. (2001). *The Bayesian Choice (second edition)*. Springer Verlag, New-York.