

Bayesian inference and convex geometry: theory, methods, and algorithms.

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

Maxwell Institute for Mathematical Sciences, Heriot-Watt University

March 2019, Paris.



Outline

- 1 Bayesian inference in imaging inverse problems
- 2 MAP estimation with Bayesian confidence regions
- 3 A decision-theoretic derivation of MAP estimation
- 4 Empirical Bayes MAP estimation with unknown regularisation parameters
- 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

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

efficiently computed by convex optimisation (Chambolle and Pock, 2016).

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

- ① it provides little information about $p(x|y)$,
- ② it is not theoretically well understood (yet),
- ③ it struggles with unknown/partially unknown models.

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

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

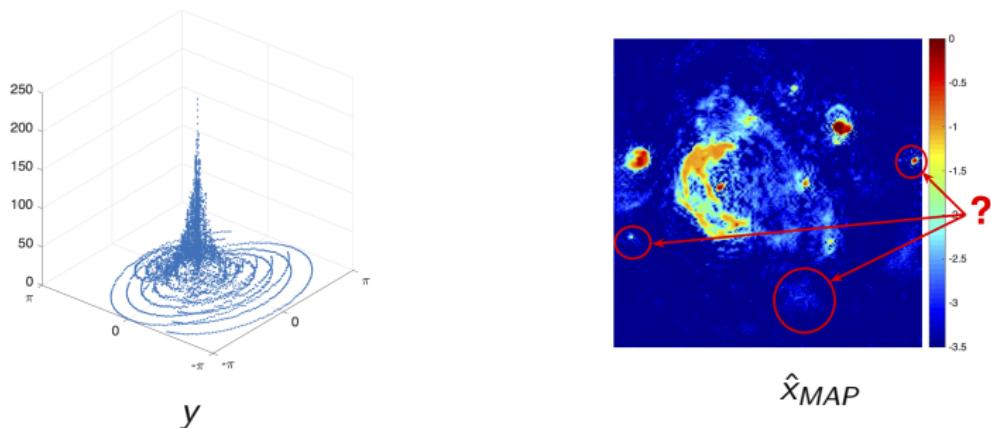


Figure: Radio-interferometric image reconstruction of the W28 supernova.

Outline

- 1 Bayesian inference in imaging inverse problems
- 2 MAP estimation with Bayesian confidence regions
- 3 A decision-theoretic derivation of MAP estimation
- 4 Empirical Bayes MAP estimation with unknown regularisation parameters
- 5 Conclusion

Posterior credible regions

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.

- We could estimate C_α^* by numerical integration (e.g., MCMC sampling), but in high-dimensional log-concave settings this is not necessary because something beautiful happens...

A concentration phenomenon...

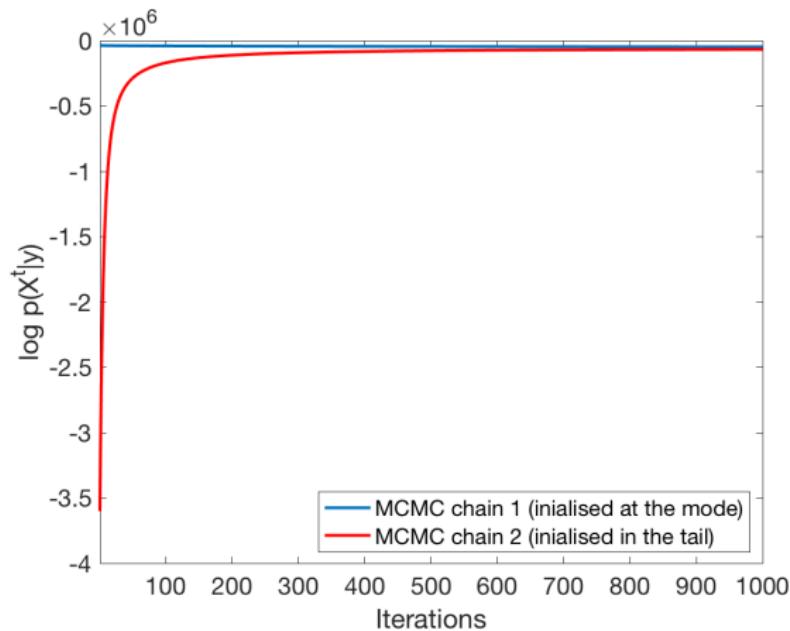


Figure: Convergence to “typical” set $\{x : \log p(x|y) \approx E[\log p(x|y)]\}$.

Proposed approximation of C_α^*

Theorem 2.1 (Pereyra (2016))

Suppose that the posterior $p(x|y) = \exp\{-\phi(x)\}/Z$ is log-concave on \mathbb{R}^d . Then, for any $\alpha \in (4 \exp(-d/3), 1)$, the HPD region C_α^* is contained by

$$\tilde{C}_\alpha = \{x : \phi(x) \leq \phi(\hat{x}_{MAP}) + \sqrt{d}\tau_\alpha + d\},$$

with universal positive constant $\tau_\alpha = \sqrt{16 \log(3/\alpha)}$ independent of $p(x|y)$.

Remark 1: \tilde{C}_α is a conservative approximation of C_α^* , i.e.,

$$x \notin \tilde{C}_\alpha \implies x \notin C_\alpha^*.$$

Remark 2: \tilde{C}_α is available as a by-product in any convex inverse problem that is solved by MAP estimation!

Approximation error bounds

Is \tilde{C}_α a reliable approximation of C_α^* ?

Theorem 2.2 (Finite-dimensional error bound (Pereyra, 2016))

Let $\tilde{\gamma}_\alpha = \phi(\hat{x}_{MAP}) + \sqrt{d}\tau_\alpha + d$. If $p(x|y)$ is log-concave on \mathbb{R}^d , then

$$0 \leq \frac{\tilde{\gamma}_\alpha - \gamma_\alpha}{d} \leq 1 + \eta_\alpha d^{-1/2},$$

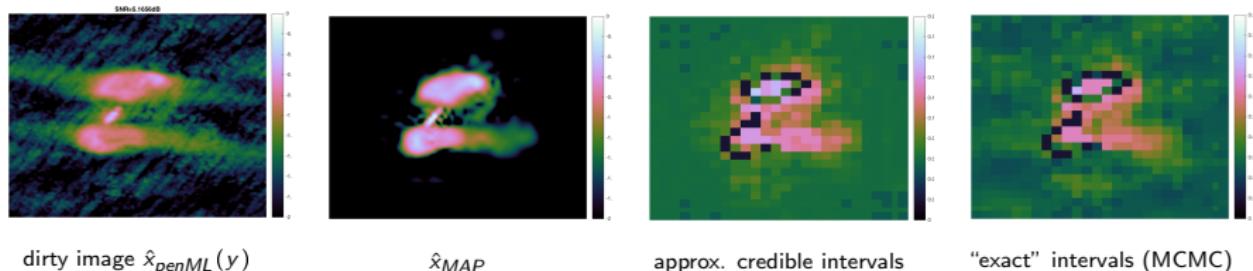
with universal positive constant $\eta_\alpha = \sqrt{16 \log(3/\alpha)} + \sqrt{1/\alpha}$.

Remark 3: \tilde{C}_α is stable (as d becomes large, the error $(\tilde{\gamma}_\alpha - \gamma_\alpha)/d \lesssim 1$).

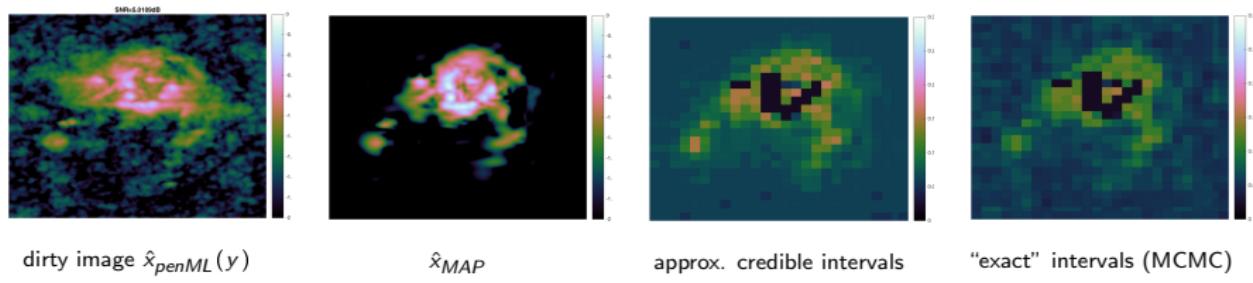
Remark 4: The lower and upper bounds are asymptotically tight w.r.t. d .

Uncertainty visualisation in radio-interferometric imaging

Astro-imaging experiment with redundant wavelet frame (Cai et al., 2017).
Local credible intervals at scale 10×10 pixels.



3C2888 radio galaxy (size 256×256 pixels, comp. time 1.8 secs.)



M31 radio galaxy (size 256×256 pixels, comp. time 1.8 secs.)

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)

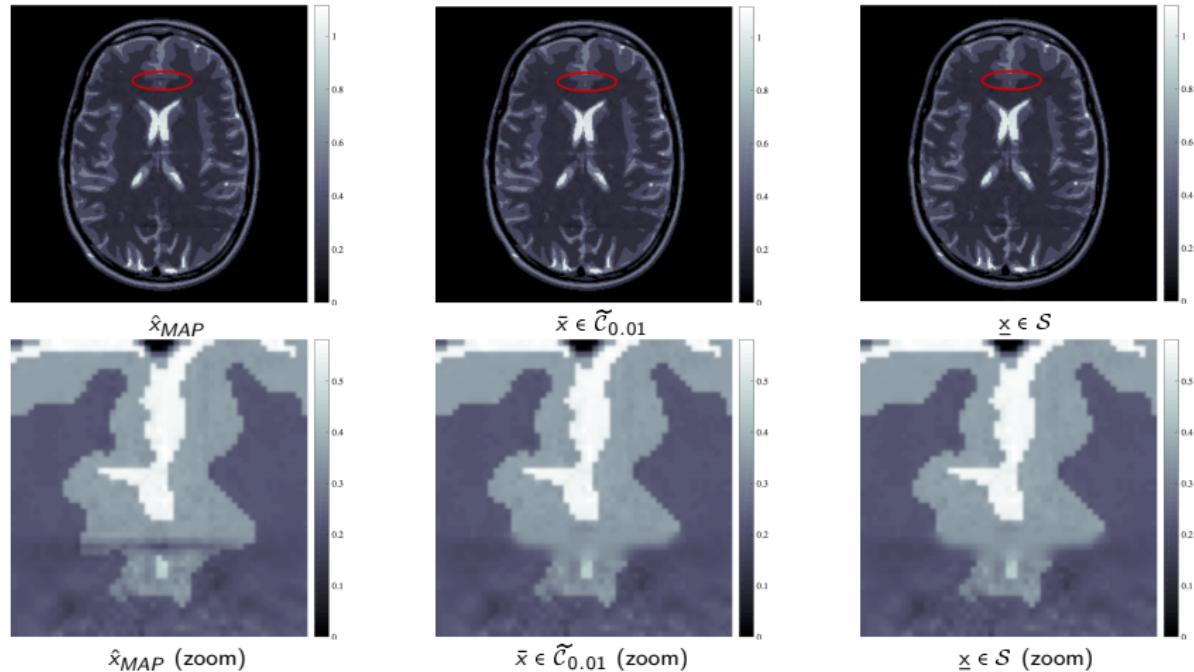
Let \mathcal{S} denote the region of \mathbb{R}^d associated with H_0 , containing all images *without the structure* of interest. Then

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

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

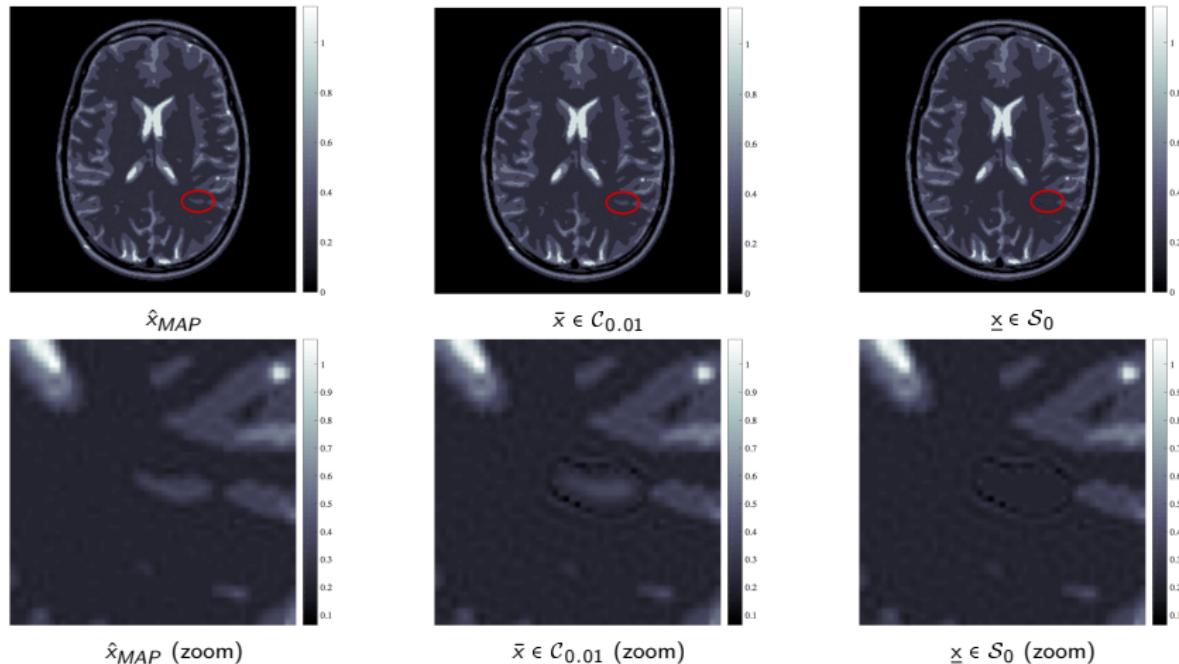
$$\min_{\bar{x}, \underline{x} \in \mathbb{R}^d} \|\bar{x} - \underline{x}\|_2^2 \quad \text{s.t.} \quad \bar{x} \in \tilde{\mathcal{C}}_\alpha, \quad \underline{x} \in \mathcal{S}.$$

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.

Outline

- 1 Bayesian inference in imaging inverse problems
- 2 MAP estimation with Bayesian confidence regions
- 3 A decision-theoretic derivation of MAP estimation
- 4 Empirical Bayes MAP estimation with unknown regularisation parameters
- 5 Conclusion

Bayesian point estimators

Bayesian point estimators arise from the decision "what point $\hat{x} \in \mathbb{R}^d$ summarises $x|y$ best?". The optimal decision under uncertainty is

$$\hat{x}_L = \operatorname{argmin}_{u \in \mathbb{R}^d} E\{L(u, x)|y\} = \operatorname{argmin}_{u \in \mathbb{R}^d} \int L(u, x)p(x|y)dx$$

where the loss $L(u, x)$ measures the "dissimilarity" between u and x .

Example: Euclidean setting $L(u, x) = \|u - x\|^2$ and $\hat{x}_L = \hat{x}_{MMSE} = E\{x|y\}$.

General desiderata:

- ① $L(u, x) \geq 0, \forall u, x \in \mathbb{R}^d,$
- ② $L(u, x) = 0 \iff u = x,$
- ③ L strictly convex w.r.t. its first argument (for estimator uniqueness).

Bayesian point estimators

Does the convex geometry of $p(x|y)$ define an interesting loss $L(u, x)$?

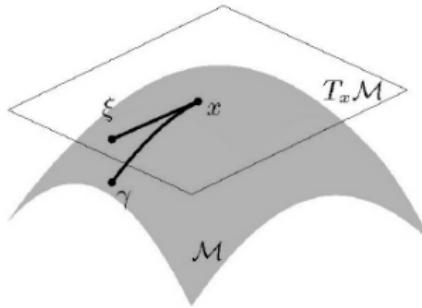
We use **differential geometry** to relate the convexity of $p(x|y)$, the geometry of the parameter space, and the loss L to perform estimation.

Differential geometry

A Riemannian manifold $\mathcal{M} = (\mathbb{R}^d, g)$, with metric $g : \mathbb{R}^d \rightarrow \mathcal{S}_{++}^d$ and global coordinate system x , is a vector space that is locally Euclidean.

For any point $x \in \mathbb{R}^d$ we have an Euclidean tangent space $T_x \mathbb{R}^d$ with inner product $\langle u, x \rangle = u^\top g(x)x$ and norm $\|x\| = \sqrt{x^\top g(x)x}$.

This geometry is local and may vary smoothly from $T_x \mathbb{R}^d$ to $T_{x'} \mathbb{R}^d$ following the affine connection $\Gamma \in \mathbb{R}^{d \times d \times d}$, given by $\Gamma_{ij,k}(x) = \partial_k g_{i,j}(x)$.



Divergence functions

Similarly to Euclidean spaces, the manifold (\mathbb{R}^d, g) supports divergences:

Definition 1 (Divergence functions)

A function $D : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a divergence function on \mathbb{R}^d if the following conditions hold for any $u, x \in \mathbb{R}^d$:

- $D(u, x) \geq 0, \forall u, x \in \mathbb{R}^d,$
- $D(u, x) = 0 \iff x = u,$
- $D(u, x)$ is strongly convex w.r.t. u , and \mathcal{C}^2 w.r.t u and x .

Canonical divergence

We focus on the *canonical divergence* on (\mathbb{R}^d, g) , a generalisation of the Euclidean squared distance to this kind of manifold:

Definition 2 (Canonical divergence (Ay and Amari, 2015))

For any $(u, x) \in \mathbb{R}^d \times \mathbb{R}^d$, the canonical divergence on (\mathbb{R}^d, g) is given by

$$D(u, x) = \int_0^1 t \dot{\gamma}_t^\top g(\gamma_t) \dot{\gamma}_t dt \quad (3)$$

where γ_t is the Γ -geodesic from u to x and $\dot{\gamma}_t = d/dt \gamma_t$.

- ① D fully specifies (\mathbb{R}^d, g) and vice-versa.
- ② $D(x + dx, x) = \|dx\|^2/2 + o(\|dx\|^2)$ where $\|\cdot\|$ is the norm on $T_x \mathbb{R}^d$.
- ③ For Euclidean space with $\langle u, x \rangle = u^\top g x$, $D(u, x) = \frac{1}{2}(u - x)^\top g(u - x)$.

Differential-geometric derivation of \hat{x}_{MAP} and \hat{x}_{MMSE}

Theorem 3 (Canonical Bayesian estimators - Part 1 (Pereyra, 2016))

Suppose that $\phi(x) = -\log p(x|y)$ is strongly convex, continuous, and \mathcal{C}^3 on \mathbb{R}^d . Let (\mathbb{R}^d, g) be the manifold induced by ϕ , i.e., $g_{i,j}(x) = \partial_i \partial_j \phi(x)$. Then, the canonical divergence on (\mathbb{R}^d, g) is the ϕ -Bregman divergence

$$D_\phi(u, x) = \phi(u) - \phi(x) - \nabla \phi(x)(u - x).$$

Remark: Because ϕ is strongly convex, then $\phi(u) > \phi(x) - \nabla \phi(x)(u - x)$ for any $u \neq x$. The divergence $D_\phi(u, x)$ quantifies this gap, related to the length of the affine geodesic from u to x on the space induced by $p(x|y)$.

Differential-geometric derivation of \hat{x}_{MAP} and \hat{x}_{MMSE}

Theorem 4 (Canonical Bayesian estimators - Part 2 (Pereyra, 2016))

The Bayesian estimator associated with $D_\phi(u, x)$ is unique and is given by

$$\begin{aligned}\hat{x}_{D_\phi} &\triangleq \underset{u \in \mathbb{R}^d}{\operatorname{argmin}} \mathbb{E}_{x|y}[D_\phi(u, x)], \\ &= \underset{x \in \mathbb{R}^d}{\operatorname{argmin}} \phi(x), \\ &= \hat{x}_{MAP}.\end{aligned}$$

Remark2: \hat{x}_{MAP} stems from Bayesian decision theory, and hence it stands on the same theoretical footing as the core Bayesian methodologies.

Remark3: The definition of the MAP estimator as the maximiser $\hat{x}_{MAP} = \operatorname{argmax}_{x \in \mathbb{R}^d} p(x|y)$ is mainly algorithmic for these models.

Differential-geometric derivation of \hat{x}_{MAP} and \hat{x}_{MMSE}

Theorem 5 (Canonical Bayesian estimators - Part 3 (Pereyra, 2016))

Moreover, the Bayesian estimator associated with the dual canonical divergence $D_\phi^*(u, x) = D_\phi(x, u)$ is also unique and is given by

$$\begin{aligned}\hat{x}_{D_\phi^*} &\triangleq \underset{u \in \mathbb{R}^d}{\operatorname{argmin}} E_{x|y}[D_\phi^*(u, x)], \\ &= \int_{\mathbb{R}^d} x p(x|y) dx, \\ &= \hat{x}_{MMSE}.\end{aligned}$$

Remark 4: \hat{x}_{MAP} and \hat{x}_{MMSE} exhibit a surprising duality, arising from the asymmetry of the canonical divergence that $p(x|y)$ induces on \mathbb{R}^d .

Remark 5: These results carry partially to models that are not strongly convex, not smooth, or that involve constraints on the parameter space.

Expected estimation error bound

Are \hat{x}_{MAP} and \hat{x}_{MMSE} “good” estimators of $x|y$?

Proposition 3.1 (Expected canonical error bound)

Suppose that $\phi(x) = -\log \pi(x|y)$ is convex on \mathbb{R}^d and \mathcal{C}^1 . Then,

$$\mathbb{E}_{x|y} [D_\phi^*(\hat{x}_{MMSE}, x)/d] \leq \mathbb{E}_{x|y} [D_\phi^*(\hat{x}_{MAP}, x)/d] \leq 1.$$

Proposition 3.2 (Expected error w.r.t. regularisation function)

Also assume that the regularisation $h(x) = -\log p(x)$ is convex. Then,

$$\mathbb{E}_{x|y} [D_h^*(\hat{x}_{MMSE}, x)/d] \leq \mathbb{E}_{x|y} [D_h^*(\hat{x}_{MAP}, x)/d] \leq 1.$$

Remark 6: These are high-dimensional stability results for \hat{x}_{MAP} and \hat{x}_{MMSE} ; the estimation error cannot grow faster than the number of pixels.

Example 1: denoising with wavelet shrinkage prior

Consider a linear problem of the form $y = Ax + w$ and a shrinkage prior on the wavelet coefficients $z = Wx$. We consider the smoothed Laplace prior

$$p(z) \propto \exp\left\{-\sum_{i=1}^d \lambda \sqrt{z_i^2 + \gamma^2}\right\}$$

where $\lambda \in \mathbb{R}^+$ and $\gamma \in \mathbb{R}^+$ are scale and shape parameters.

The likelihood is $p(y|z) \propto \exp\left\{-\frac{1}{2\sigma^2} \|y - AW^\top z\|_2^2\right\}$ and hence

$$p(z|y) \propto \exp\left\{-\frac{1}{2\sigma^2} \|y - AW^\top z\|_2^2 - \sum_{i=1}^d \lambda \sqrt{z_i^2 + \gamma^2}\right\}$$

This model is \mathcal{C}^∞ and strongly log-concave, and hence the theory applies.

Example 1: denoising with wavelet shrinkage prior

To analyse the geometry induced by $\phi(z) = -\log p(z|y)$ we suppose that $A = \mathbb{I}$ and $W^\top W = \mathbb{I}$, and obtain $D_\phi(u, z) = \sum_{i=1}^d D_\psi(u_i, z_i)$ with

$$D_\psi(u_i, z_i) = \frac{1}{2\sigma^2}(u_i - z_i)^2 + \lambda \frac{\sqrt{z_i^2 + \gamma^2} \sqrt{u_i^2 + \gamma^2} - z_i u_i - \gamma^2}{\sqrt{z_i^2 + \gamma^2}}.$$

The non-quadratic term introduces **additional shrinkage** and leads to the differences between x_{MMSE} and x_{MAP} .

Example 1: denoising with wavelet shrinkage prior

Illustration with the Flinstones image ($\sigma = 0.08$, $\lambda = 12$ and $\gamma = 0.01$).



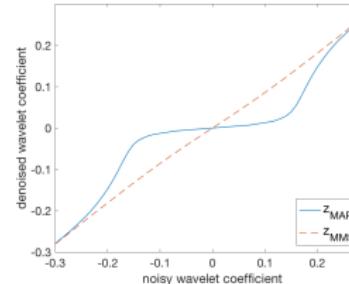
noisy image y (SNR 17.6dB)



\hat{x}_{MAP} (SNR 19.8dB)



\hat{x}_{MMSE} (SNR 17.7dB)

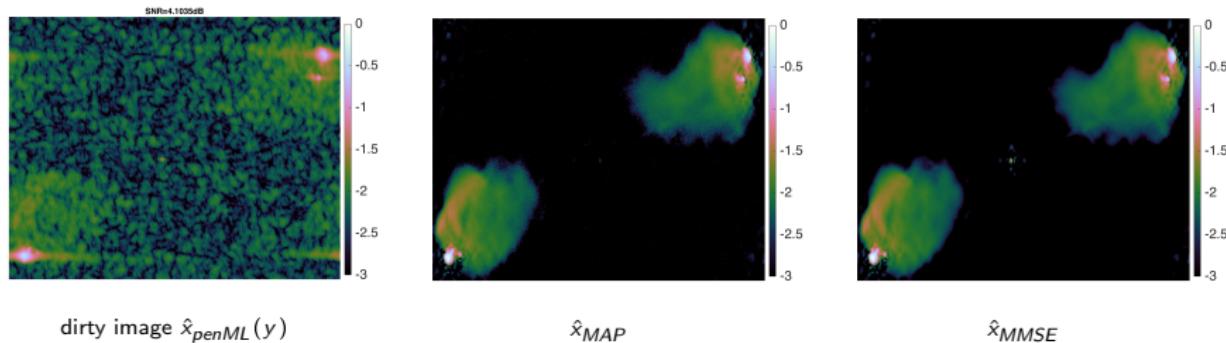


denoising functions for \hat{x}_{MAP} and \hat{x}_{MMSE}

Illustrative example of a model where the action of the shrinkage prior acts predominantly via D_ψ (Note: setting $\gamma = 0$ leads to \hat{x}_{MAP} with SNR 18.8dB).

Illustrative example: astronomical image reconstruction

Generalisation warning: shrinkage priors can also act predominantly via the model (not D_ψ), producing similar \hat{x}_{MAP} and \hat{x}_{MMSE} results; e.g.,



Radio-interferometric imaging of the Cygnus A galaxy Cai et al. (2017).

Outline

- 1 Bayesian inference in imaging inverse problems
- 2 MAP estimation with Bayesian confidence regions
- 3 A decision-theoretic derivation of MAP estimation
- 4 Empirical Bayes MAP estimation with unknown regularisation parameters
- 5 Conclusion

Problem statement

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

When 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} at rate $O(1/m)$, with poss. acceleration to $O(1/m^2)$ ¹.

However, θ is generally unknown, significantly complicating the problem.

¹Recall that the proximal operator $\operatorname{prox}_{\varphi}^{\lambda}(x) \triangleq \operatorname{argmin}_{u \in \mathbb{R}^N} \varphi(u) + \frac{1}{2\lambda} \|u - x\|^2$.

Regularisation parameter MLE

In this talk we adopt an empirical Bayes approach and consider the MLE

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

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

$$\hat{x}_{MAP} = \operatorname{argmin}_{x \in \mathbb{R}^d} f_y(x) + \hat{\theta} \varphi(x). \quad (4)$$

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

$$\theta = P_{\Theta} \left(\theta + \delta E_{x|\theta} \{ \varphi(x) \} - \delta E_{x|y,\theta} \{ \varphi(x) \} \right), \quad (5)$$

which we solve by using a stochastic approximation algorithm.

SAPG algorithm driven by MCMC kernels

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 MCMC² updates: generate $z^{(t+1)} \sim \mathcal{N}(0, \mathbb{I}_d)$

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

$$u^{(t+1)} = \left(1 - \frac{\gamma}{\lambda}\right)u^{(t)} + \frac{\gamma}{\lambda} \text{prox}_{\varphi}^{\theta\lambda}\left(u^{(t)}\right) + \sqrt{2\gamma} z^{(t+1)},$$

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 $\hat{\theta}$.

²Langevin MCMC kernels are highly efficient and scale as $\mathcal{O}(d \log(d))$, see Pereyra (2015); Durmus et al. (2018).

Illustrative example - Image deblurring with ℓ_1 prior

We consider the Bayesian image deblurring model

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

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

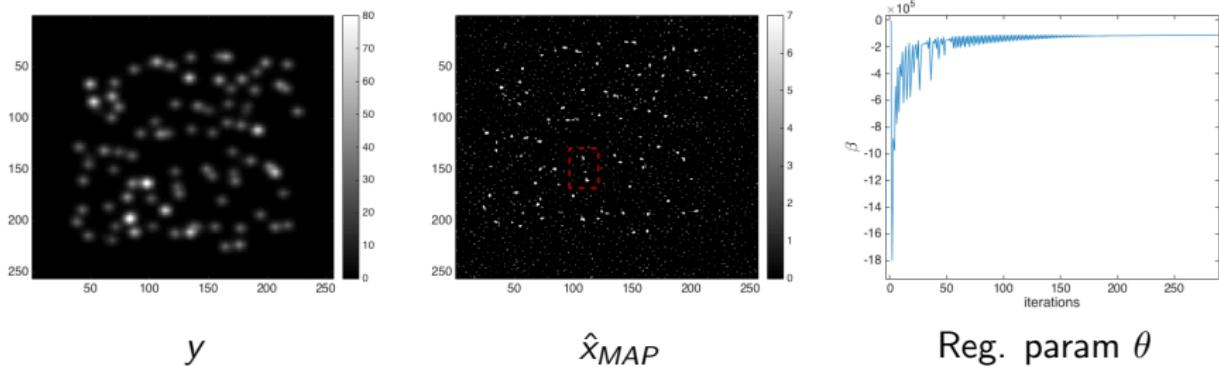
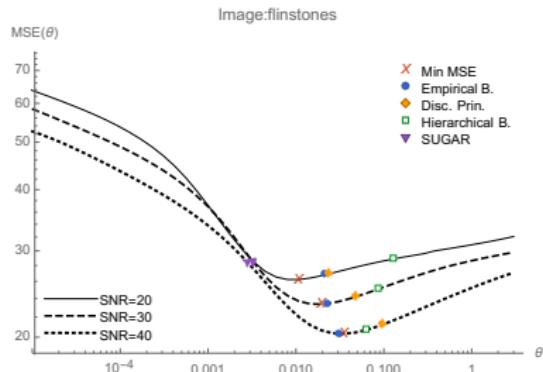
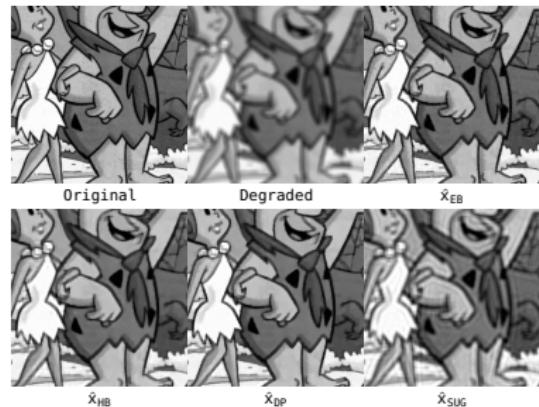


Figure: Molecules image deconvolution experiment, computing time 0.75 secs.

Deblurring with Total-Variation Prior

	SNR=20dB		SNR=30		SNR=40	
	MSE	Time (min)	MSE	Time (min)	MSE	Time (min)
Best	23.29		21.39		19.06	
EB	23.52	10.33	21.47	10.13	19.21	9.49
HB	25.07	0.58	22.84	1.27	19.84	3.27
DP	23.73		21.87		19.78	
SUG	24.44	3.92	24.24	4.50	24.21	4.81



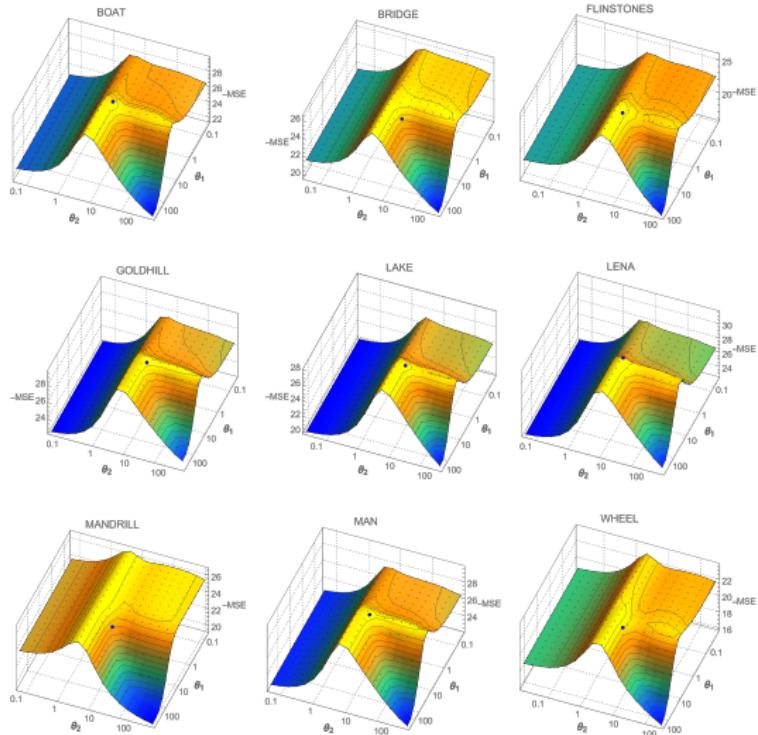
Denoising with Total Generalized Variation

We consider $TGV_\theta^2(u) = \inf_{v \in BD(\Omega)} \theta_1 \int_{\Omega} |\nabla u - v| + \theta_2 \int_{\Omega} |\varepsilon(v)|$ with $k = 2$.

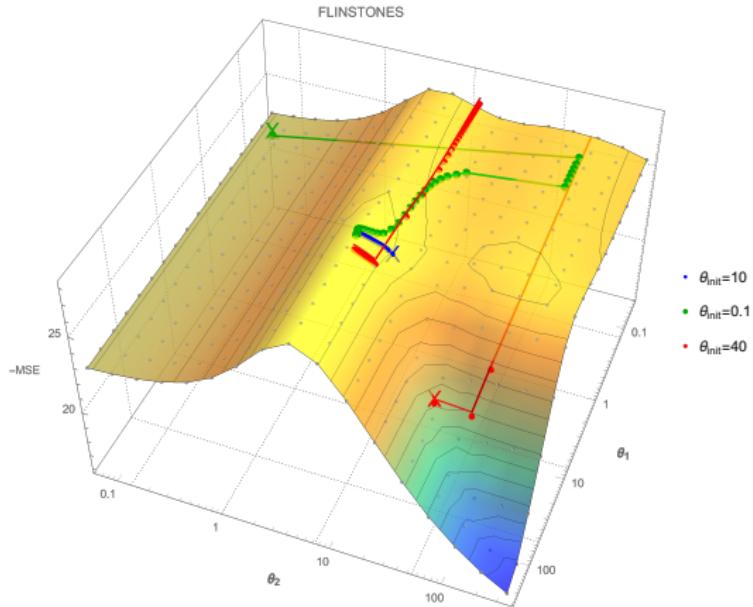


Figure: *Goldhill* image (Original-Degraded-Estimated MAP), SNR=12dB.

Denoising with Total Generalized Variation

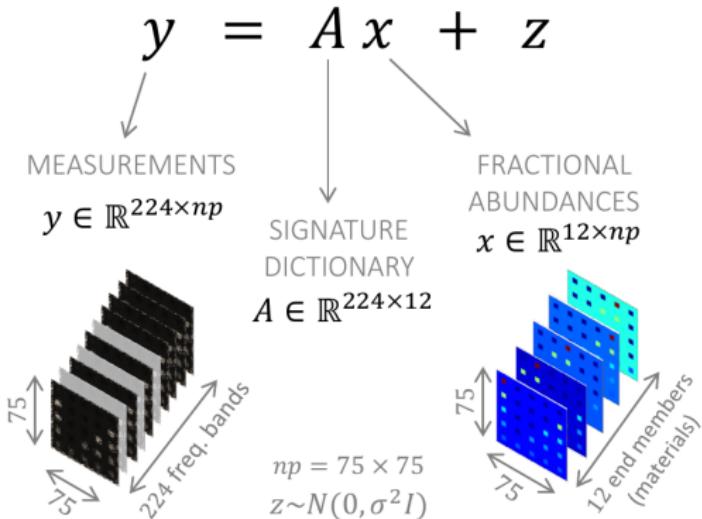


Evolution of θ through iterations starting from different initial values:



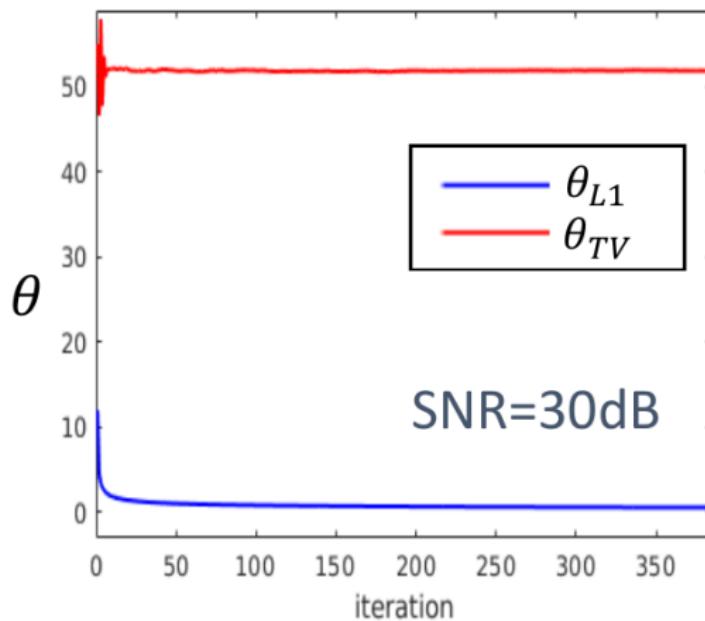
Hyperspectral Unmixing

- Objective: to recover fractional abundances x from the mixed noisy spectral signatures y measured for every pixel.

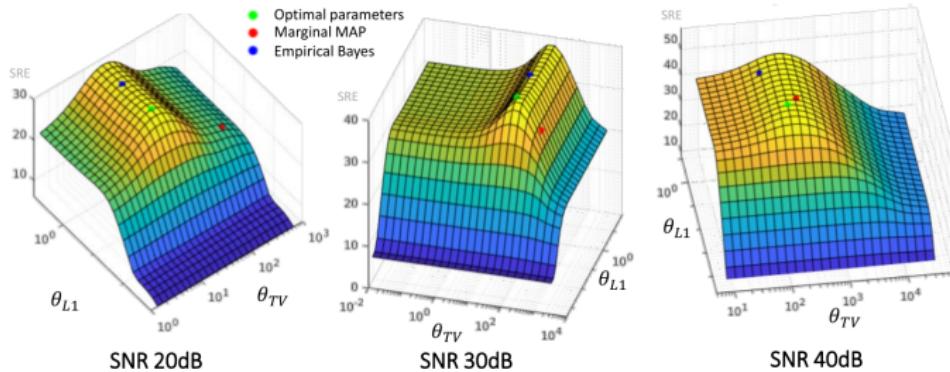


- Regularizer: $\varphi(x) = \theta_{TV} TV(x) + \theta_{L1} \|x\|_1$ s.t. $x \geq 0$
- We estimate θ_{TV} and θ_{L1} for synthetic test images

- Evolution of θ_{TV} and θ_{L1} with iterations.



- SRE plots for different SNR values



Outline

- 1 Bayesian inference in imaging inverse problems
- 2 MAP estimation with Bayesian confidence regions
- 3 A decision-theoretic derivation of MAP estimation
- 4 Empirical Bayes MAP estimation with unknown regularisation parameters
- 5 Conclusion

Conclusion

- The challenges facing modern imaging sciences require a methodological paradigm shift to go beyond point estimation.
- Great potential for synergy between Bayesian and variational approaches at algorithmic, methodological, and theoretical levels.

Thank you!



We are hiring: **4 permanent positions in probability and statistical data science!**

Bibliography:

- Ay, N. and Amari, S.-I. (2015). A novel approach to canonical divergences within information geometry. *Entropy*, 17(12):7866.
- 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.
- 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*.
- 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>.
- Pereyra, M. (2016). Maximum-a-posteriori estimation with bayesian confidence regions. *SIAM J. Imaging Sci.*, 6(3):1665–1688.
- Pereyra, M. (2016). Revisiting maximum-a-posteriori estimation in log-concave models: from differential geometry to decision theory. *ArXiv e-prints*.
- 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.