Scalable Bayesian uncertainty quantification with learned convex regularisers

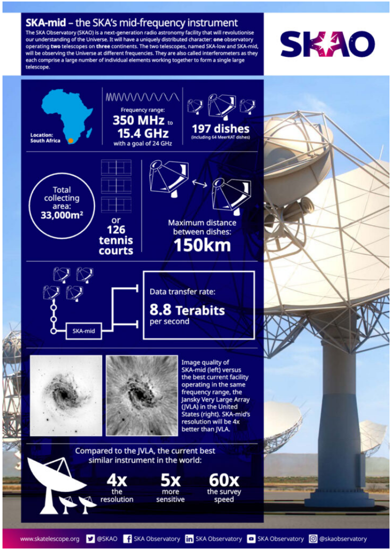
Tobías I. Liaudat
Computer Science department, University College London

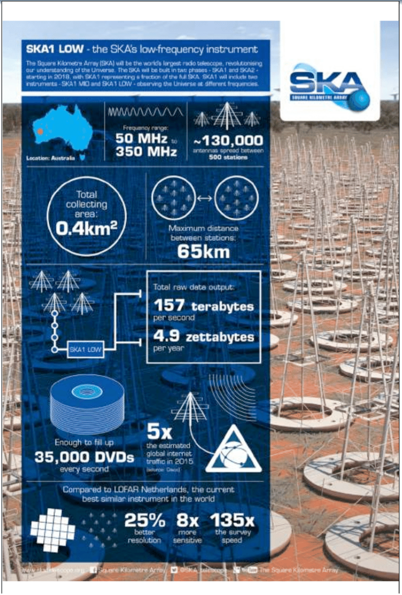
In collabortaion with Jason D. McEwen, Marta Betcke and Marcelo Pereyra

CMIC-WEISS Joint seminar series

14th June 2023

Motivation: SKA's radio interferometer





Radio interferometric imaging

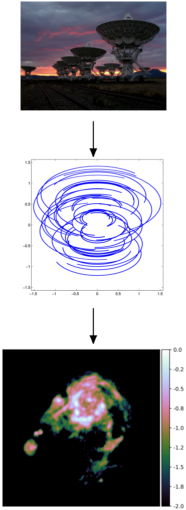
Linear observational model

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}$$

- $\mathbf{y} \in \mathbb{C}^M$: Observed Fourier coefficients
- $\mathbf{n} \in \mathbb{C}^M$: Observational noise (White and Gaussian)
- $\mathbf{x} \in \mathbb{R}^N$: Sky intensity image
- $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$: Linear measurement operator
  - FFT and Fourier mask

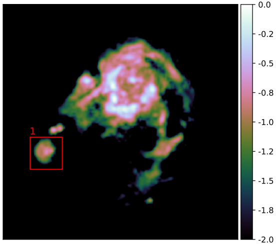
Due to $\mathbf{n}$ and $\mathbf{\Phi}$ the inverse problem is ill-posed

We need to estimate $\hat{\mathbf{x}}$ from $\mathbf{y}$



$\hat{\mathbf{x}}$

Image reconstruction: $\hat{\mathbf{x}}$



Is this blob *physical*?

→ Is it a reconstruction artefact?

→ Is it backed by the data?

Several reasons to develop uncertainty quantification (UQ) techniques for the reconstruction

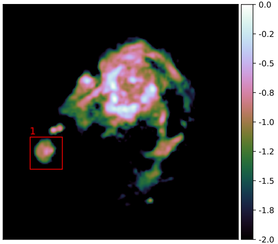
Usual UQ techniques from the Bayesian framework rely on interrogating the posterior exploiting Bayes' theorem:

$$\underbrace{p(\mathbf{x}|\mathbf{y})}_{\text{Posterior}} \propto \underbrace{p(\mathbf{y}|\mathbf{x})}_{\text{Likelihood}} \underbrace{p(\mathbf{x})}_{\text{Prior}}$$

Represent the posterior through samples drawn from $\sim p(\mathbf{x}|\mathbf{y})$ obtained through Markov chain Monte Carlo (MCMC)

For example, Cai et al. (2018) applies this for radio imaging

Image reconstruction: $\hat{\mathbf{x}}$



Is this blob *physical*?

→ Is it a reconstruction artefact?

→ Is it backed by the data?

Several reasons to develop uncertainty quantification (UQ) techniques for the reconstruction

Usual UQ techniques from the Bayesian framework rely on interrogating the posterior exploiting Bayes' theorem:

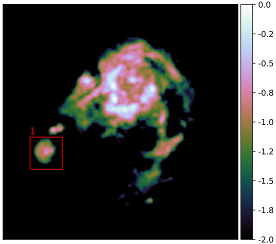
$$\underbrace{p(\mathbf{x}|\mathbf{y})}_{\text{Posterior}} \propto \underbrace{p(\mathbf{y}|\mathbf{x})}_{\text{Likelihood}} \underbrace{p(\mathbf{x})}_{\text{Prior}}$$

Represent the posterior through samples drawn from $\sim p(\mathbf{x}|\mathbf{y})$ obtained through Markov chain Monte Carlo (MCMC)

For example, Cai et al. (2018) applies this for radio imaging

Uncertainty quantification: more than a point estimate

Image reconstruction: $\hat{\mathbf{x}}$



Is this blob *physical*?

→ Is it a reconstruction artefact?

→ Is it backed by the data?

Several reasons to develop uncertainty quantification (UQ) techniques for the reconstruction

Usual UQ techniques from the Bayesian framework rely on interrogating the posterior exploiting Bayes' theorem:

$$\underbrace{p(\mathbf{x}|\mathbf{y})}_{\text{Posterior}} \propto \underbrace{p(\mathbf{y}|\mathbf{x})}_{\text{Likelihood}} \underbrace{p(\mathbf{x})}_{\text{Prior}}$$

Represent the posterior through samples drawn from $\sim p(\mathbf{x}|\mathbf{y})$ obtained through Markov chain Monte Carlo (MCMC)

For example, Cai et al. (2018) applies this for radio imaging

Uncertainty quantification: more than a point estimate

Cai et al. (2018) approach:

1. Define a likelihood $p(\mathbf{y}|\mathbf{x}) = \exp[-f(\mathbf{x}, \mathbf{y})]$
   → The Gaussian likelihood $f(\mathbf{x}, \mathbf{y})$ is known: $\|\mathbf{y} - \mathbf{\Phi x}\|_2^2 / 2\sigma^2$
2. Define a prior $p(\mathbf{x}) = \exp[-g(\mathbf{x})]$
   → Solution $\mathbf{x}$ is sparse in a wavelet dictionary $\mathbf{\Psi}$. The prior $g(\mathbf{x})$ is: $\lambda\|\mathbf{\Psi}^\dagger x\|_1$
3. Choose a point estimate
   → Use the Maximum-a-posteriori (MAP) estimation:

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\max_{\mathbf{x}\in\mathbb{R}^N} p(\mathbf{x}|\mathbf{y}) = \arg\min_{\mathbf{x}\in\mathbb{R}^N} - \log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x}),$$

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{y} - \mathbf{\Phi x}\|_2^2 / 2\sigma^2 + \lambda\|\mathbf{\Psi}^\dagger x\|_1 .$$

   → Estimate $\hat{\mathbf{x}}_{\text{MAP}}$ through **convex optimisation** using a **proximal algorithm**
4. Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^K$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$
   → **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

Uncertainty quantification: more than a point estimate

Cai et al. (2018) approach:

1. Define a likelihood $p(\mathbf{y}|\mathbf{x}) = \exp[-f(\mathbf{x}, \mathbf{y})]$
   → The Gaussian likelihood $f(\mathbf{x}, \mathbf{y})$ is known: $\|\mathbf{y} - \mathbf{\Phi x}\|_2^2 / 2\sigma^2$
2. Define a prior $p(\mathbf{x}) = \exp[-g(\mathbf{x})]$
   → Solution $\mathbf{x}$ is sparse in a wavelet dictionary $\mathbf{\Psi}$. The prior $g(\mathbf{x})$ is: $\lambda \|\mathbf{\Psi}^\dagger x\|_1$
3. Choose a point estimate
   → Use the Maximum-a-posteriori (MAP) estimation:
   $$\hat{\mathbf{x}}_{\text{MAP}} = \underset{\mathbf{x} \in \mathbb{R}^N}{\arg\max}\, p(\mathbf{x}|\mathbf{y}) = \underset{\mathbf{x} \in \mathbb{R}^N}{\arg\min} - \log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x}),$$
   $$\hat{\mathbf{x}}_{\text{MAP}} = \underset{\mathbf{x} \in \mathbb{R}^N}{\arg\min}\, \|\mathbf{y} - \mathbf{\Phi x}\|_2^2 / 2\sigma^2 + \lambda \|\mathbf{\Psi}^\dagger x\|_1 .$$
   → Estimate $\hat{\mathbf{x}}_{\text{MAP}}$ through **convex optimisation** using a **proximal algorithm**
4. Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^K$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$
   → **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

Uncertainty quantification: more than a point estimate

Cai et al. (2018) approach:

1. Define a likelihood $p(\mathbf{y}|\mathbf{x}) = \exp[-f(\mathbf{x},\mathbf{y})]$
   → The Gaussian likelihood $f(\mathbf{x},\mathbf{y})$ is known: $\|\mathbf{y} - \mathbf{\Phi x}\|_2^2/2\sigma^2$
2. Define a prior $p(\mathbf{x}) = \exp[-g(\mathbf{x})]$
   → Solution $\mathbf{x}$ is sparse in a wavelet dictionary $\mathbf{\Psi}$. The prior $g(\mathbf{x})$ is: $\lambda\|\mathbf{\Psi}^\dagger x\|_1$
3. Choose a point estimate
   → Use the Maximum-a-posteriori (MAP) estimation:

$$\hat{\mathbf{x}}_{\text{MAP}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\max}\, p(\mathbf{x}|\mathbf{y}) = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min} - \log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x}),$$

$$\hat{\mathbf{x}}_{\text{MAP}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min}\, \|\mathbf{y} - \mathbf{\Phi x}\|_2^2/2\sigma^2 + \lambda\|\mathbf{\Psi}^\dagger x\|_1.$$

   → Estimate $\hat{\mathbf{x}}_{\text{MAP}}$ through **convex optimisation** using a **proximal algorithm**
4. Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^K$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$
   → **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

Uncertainty quantification: more than a point estimate

Cai et al. (2018) approach:

1. Define a likelihood $p(\mathbf{y}|\mathbf{x}) = \exp[-f(\mathbf{x},\mathbf{y})]$
   → The Gaussian likelihood $f(\mathbf{x},\mathbf{y})$ is known: $\|\mathbf{y} - \mathbf{\Phi x}\|_2^2/2\sigma^2$
2. Define a prior $p(\mathbf{x}) = \exp[-g(\mathbf{x})]$
   → Solution $\mathbf{x}$ is sparse in a wavelet dictionary $\mathbf{\Psi}$. The prior $g(\mathbf{x})$ is: $\lambda\|\mathbf{\Psi}^\dagger x\|_1$
3. Choose a point estimate
   → Use the Maximum-a-posteriori (MAP) estimation:

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\max}\, p(\mathbf{x}|\mathbf{y}) = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min} -\log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x}),$$

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min}\, \|\mathbf{y} - \mathbf{\Phi x}\|_2^2/2\sigma^2 + \lambda\|\mathbf{\Psi}^\dagger x\|_1.$$

   → Estimate $\hat{\mathbf{x}}_{\mathrm{MAP}}$ through **convex optimisation** using a **proximal algorithm**
4. Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^{K}$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$
   → **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

Uncertainty quantification: more than a point estimate

Cai et al. (2018) approach:

1. Define a likelihood $p(\mathbf{y}|\mathbf{x}) = \exp[-f(\mathbf{x},\mathbf{y})]$
   $\rightarrow$ The Gaussian likelihood $f(\mathbf{x},\mathbf{y})$ is known: $\|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2/2\sigma^2$
2. Define a prior $p(\mathbf{x}) = \exp[-g(\mathbf{x})]$
   $\rightarrow$ Solution $\mathbf{x}$ is sparse in a wavelet dictionary $\boldsymbol{\Psi}$. The prior $g(\mathbf{x})$ is: $\lambda\|\boldsymbol{\Psi}^\dagger x\|_1$
3. Choose a point estimate
   $\rightarrow$ Use the Maximum-a-posteriori (MAP) estimation:

$$\hat{\mathbf{x}}_{\text{MAP}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\max}\, p(\mathbf{x}|\mathbf{y}) = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min} - \log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x}),$$

$$\hat{\mathbf{x}}_{\text{MAP}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min}\, \|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}\|_2^2/2\sigma^2 + \lambda\|\boldsymbol{\Psi}^\dagger x\|_1.$$

   $\rightarrow$ Estimate $\hat{\mathbf{x}}_{\text{MAP}}$ through **convex optimisation** using a **proximal algorithm**
4. Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^{K}$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$
   $\rightarrow$ **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

Uncertainty quantification: more than a point estimate

Cai et al. (2018) approach:

1. Define a likelihood $p(\mathbf{y}|\mathbf{x}) = \exp[-f(\mathbf{x}, \mathbf{y})]$
   → The Gaussian likelihood $f(\mathbf{x}, \mathbf{y})$ is known: $\|\mathbf{y} - \mathbf{\Phi x}\|_2^2 / 2\sigma^2$
2. Define a prior $p(\mathbf{x}) = \exp[-g(\mathbf{x})]$
   → Solution $\mathbf{x}$ is sparse in a wavelet dictionary $\mathbf{\Psi}$. The prior $g(\mathbf{x})$ is: $\lambda\|\mathbf{\Psi}^\dagger x\|_1$
3. Choose a point estimate
   → Use the Maximum-a-posteriori (MAP) estimation:
   $$\hat{\mathbf{x}}_{\text{MAP}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\max}\, p(\mathbf{x}|\mathbf{y}) = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min} - \log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x}),$$
   $$\hat{\mathbf{x}}_{\text{MAP}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min}\, \|\mathbf{y} - \mathbf{\Phi x}\|_2^2 / 2\sigma^2 + \lambda\|\mathbf{\Psi}^\dagger x\|_1.$$
   → Estimate $\hat{\mathbf{x}}_{\text{MAP}}$ through **convex optimisation** using a **proximal algorithm**
4. Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^K$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$
   → **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

The problem is not solved

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\mathbf{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
$\rightarrow$ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

The problem is not solved

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\mathbf{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
$\rightarrow$ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

The problem is not solved

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\mathbf{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
$\rightarrow$ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\mathbf{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
$\rightarrow$ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

The problem is not solved

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\mathbf{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
$\rightarrow$ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

The problem is not solved

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\boldsymbol{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
$\rightarrow$ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

The problem is not solved

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\boldsymbol{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
$\rightarrow$ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

Highest posterior density region

Posterior credible region:

$$p(\mathbf{x} \in C_\alpha | \mathbf{y}) = \int_{\mathbf{x} \in \mathbb{R}^N} p(\mathbf{x}|\mathbf{y}) \mathbb{1}_{C_\alpha} \mathrm{d}\mathbf{x} = 1 - \alpha,$$

We consider the highest posterior density (HPD) region

$$C^*_\alpha = \{\mathbf{x} : \underbrace{f(\mathbf{x}) + g(\mathbf{x})}_{\text{potential}} \leq \gamma_\alpha\}, \quad \text{with } \gamma_\alpha \in \mathbb{R}, \quad \text{and } p(\mathbf{x} \in C^*_\alpha | \mathbf{y}) = 1 - \alpha \text{ holds,}$$

Theorem 3.1 (Pereyra, 2017)

Suppose the posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z$ is log-concave on $\mathbb{R}^N$. Then, for any $\alpha \in (4\exp[(-N/3)], 1)$, the HPD region C^*_α is contained by

$$\hat{C}_\alpha = \left\{\mathbf{x} : f(\mathbf{x}) + g(\mathbf{x}) \leq \hat{\gamma}_\alpha = f(\hat{\mathbf{x}}_{\text{MAP}}) + g(\hat{\mathbf{x}}_{\text{MAP}}) + \sqrt{N}\tau_\alpha + N\right\},$$

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

We only need to evaluate $f + g$ on the MAP estimation $\hat{\mathbf{x}}_{\text{MAP}}$!

Highest posterior density region

Posterior credible region:

$$p(\mathbf{x} \in C_\alpha | \mathbf{y}) = \int_{\mathbf{x} \in \mathbb{R}^N} p(\mathbf{x}|\mathbf{y}) \mathbb{1}_{C_\alpha} \mathrm{d}\mathbf{x} = 1 - \alpha,$$

We consider the highest posterior density (HPD) region

$$C^*_\alpha = \{\mathbf{x} : \underbrace{f(\mathbf{x}) + g(\mathbf{x})}_{\text{potential}} \leq \gamma_\alpha\}, \quad \text{with } \gamma_\alpha \in \mathbb{R}, \quad \text{and } p(\mathbf{x} \in C^*_\alpha|\mathbf{y}) = 1 - \alpha \text{ holds},$$

Theorem 3.1 (Pereyra, 2017)

Suppose the posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z$ is log-concave on $\mathbb{R}^N$. Then, for any $\alpha \in (4\exp[(-N/3)], 1)$, the HPD region C^*_α is contained by

$$\hat{C}_\alpha = \left\{\mathbf{x} : f(\mathbf{x}) + g(\mathbf{x}) \leq \hat{\gamma}_\alpha = f(\hat{\mathbf{x}}_{\mathrm{MAP}}) + g(\hat{\mathbf{x}}_{\mathrm{MAP}}) + \sqrt{N}\tau_\alpha + N\right\},$$

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

We only need to evaluate $f + g$ on the MAP estimation $\hat{\mathbf{x}}_{\mathrm{MAP}}$!

Highest posterior density region

Posterior credible region:

$$p(\mathbf{x} \in C_\alpha | \mathbf{y}) = \int_{\mathbf{x} \in \mathbb{R}^N} p(\mathbf{x}|\mathbf{y}) \mathbb{1}_{C_\alpha} \mathrm{d}\mathbf{x} = 1 - \alpha,$$

We consider the highest posterior density (HPD) region

$$C^*_\alpha = \{\mathbf{x} : \underbrace{f(\mathbf{x}) + g(\mathbf{x})}_{\text{potential}} \leq \gamma_\alpha\}, \quad \text{with } \gamma_\alpha \in \mathbb{R}, \quad \text{and } p(\mathbf{x} \in C^*_\alpha | \mathbf{y}) = 1 - \alpha \text{ holds},$$

Theorem 3.1 (Pereyra, 2017)

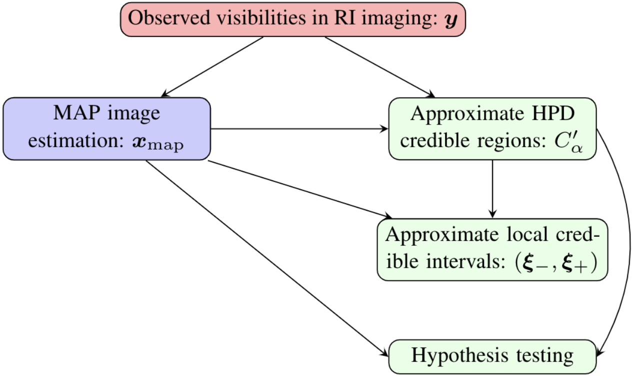
Suppose the posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z$ is log-concave on $\mathbb{R}^N$. Then, for any $\alpha \in (4\exp[(-N/3)], 1)$, the HPD region C^*_α is contained by

$$\hat{C}_\alpha = \left\{\mathbf{x} : f(\mathbf{x}) + g(\mathbf{x}) \leq \hat{\gamma}_\alpha = f(\hat{\mathbf{x}}_{\text{MAP}}) + g(\hat{\mathbf{x}}_{\text{MAP}}) + \sqrt{N}\tau_\alpha + N\right\},$$

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

We only need to evaluate $f + g$ on the MAP estimation $\hat{\mathbf{x}}_{\text{MAP}}$!

MAP-based uncertainty quantification



Hypothesis test with significance α:

1. Calculate the MAP: $\mathbf{x}_{\text{MAP}}$
2. Compute HPD region threshold $\hat{\gamma}_\alpha$
3. Construct a surrogate image $\mathbf{x}_{\text{sgt}}$
4. Compute $\mathcal{E} = f(\mathbf{x}_{\text{sgt}}) + g(\mathbf{x}_{\text{sgt}})$
5. If $\mathcal{E} \leq \hat{\gamma}_\alpha \rightarrow$ inconclusive test
6. If $\mathcal{E} > \hat{\gamma}_\alpha \rightarrow$ reject hypothesis

Pixel-level UQ visualisation

Local credible intervals (LCI)

Scalable Bayesian uncertainty quantification

1. **Scalability** → Need to rely on optimisation ~~sampling~~, use the MAP estimator
2. **Uncertainty quantification** → Need the potential to be convex and explicit
3. **Good reconstruction** → Need to use data-driven (learned) approaches

The approach requires our prior to be convex and with an explicit potential

We constrain our prior to be convex, but we gain an effortless UQ!

Scalable Bayesian uncertainty quantification

1. **Scalability** → Need to rely on optimisation ~~sampling~~, use the MAP estimator
2. **Uncertainty quantification** → Need the potential to be convex and explicit
3. **Good reconstruction** → Need to use data-driven (learned) approaches

The approach requires our prior to be convex and with an explicit potential

We constrain our prior to be convex, but we gain an effortless UQ!

Scalable Bayesian uncertainty quantification

1. **Scalability** → Need to rely on optimisation ~~sampling~~, use the MAP estimator
2. **Uncertainty quantification** → Need the potential to be convex and explicit
3. **Good reconstruction** → Need to use data-driven (learned) approaches

The approach requires our prior to be convex and with an explicit potential

We constrain our prior to be convex, but we gain an effortless UQ!

Scalable Bayesian uncertainty quantification

1. **Scalability** $\rightarrow$ Need to rely on optimisation ~~sampling~~, use the MAP estimator
2. **Uncertainty quantification** $\rightarrow$ Need the potential to be convex and explicit
3. **Good reconstruction** $\rightarrow$ Need to use data-driven (learned) approaches

The approach requires our prior to be convex and with an explicit potential

We constrain our prior to be convex, but we gain an effortless UQ!

Scalable Bayesian uncertainty quantification

1. **Scalability** → Need to rely on optimisation ~~sampling~~, use the MAP estimator
2. **Uncertainty quantification** → Need the potential to be convex and explicit
3. **Good reconstruction** → Need to use data-driven (learned) approaches

The approach requires our prior to be convex and with an explicit potential

We constrain our prior to be convex, but we gain an effortless UQ!

Learned convex regulariser

We use the convex ridge regulariser R from Goujon et al. (2022), where

$$R : \mathbb{R}^N \mapsto \mathbb{R}, \quad R(\mathbf{x}) = \sum_{n=1}^{N_C} \sum_k \psi_n \left((\mathbf{h}_n * \mathbf{x}) [k] \right),$$

- ψ_n are learned convex profile functions with Lipschitz continuous derivate
- There are N_C learned convolutional filters $\mathbf{h}_n$
- R is trained as a (multi-)gradient step denoiser

Properties:

1. **Explicit cost**
2. **Convex**
3. **Smooth regulariser with known Lipschitz constant**

Learned convex regulariser

We use the convex ridge regulariser R from Goujon et al. (2022), where

$$R : \mathbb{R}^N \mapsto \mathbb{R}, \quad R(\mathbf{x}) = \sum_{n=1}^{N_C} \sum_{k} \psi_n \left((\mathbf{h}_n * \mathbf{x}) [k] \right),$$

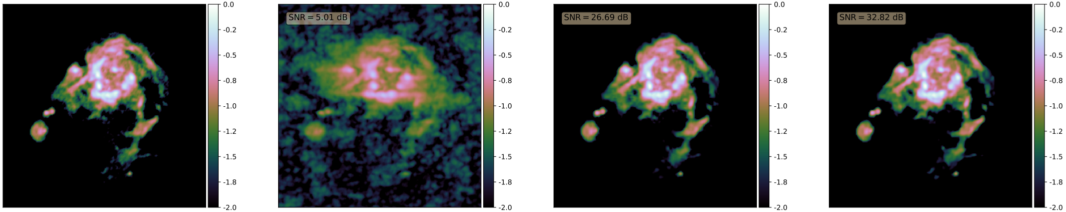
- ψ_n are learned convex profile functions with Lipschitz continuous derivate
- There are N_C learned convolutional filters $\mathbf{h}_n$
- R is trained as a (multi-)gradient step denoiser

Properties:

1. **Explicit cost**
2. **Convex**
3. **Smooth regulariser with known Lipschitz constant**

MAP reconstructions

MAP estimations were computed using the Forward-Backward splitting algorithm

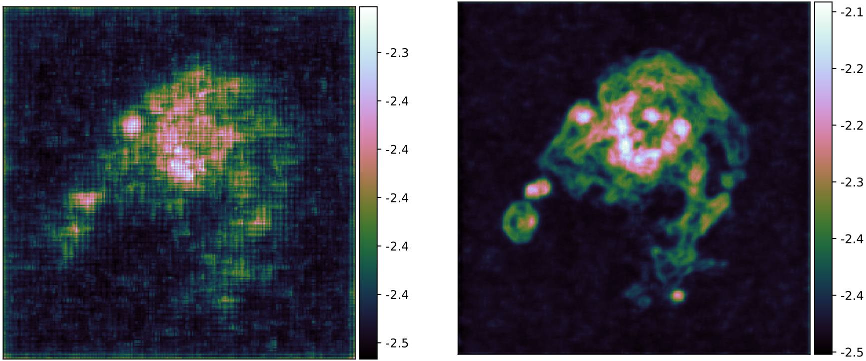


Ground truth	Dirty image	Wavelet prior	Learned prior
	SNR = 5.01 dB	SNR = 26.69 dB	**SNR = 32.82 dB**

Improved the reconstruction by 6 dB!

Posterior standard deviation

Computed using 10^4 samples obtained from the sampling algorithm SK-ROCK (Pereyra et al., 2020)



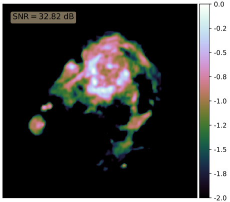
Wavelet

Learned regulariser

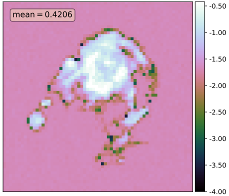
Improved quality of the posterior St Dev

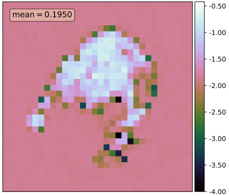
The learned convex regulariser was trained on natural images not RI images

The local credible intervals (LCI) give a local measure of uncertainty

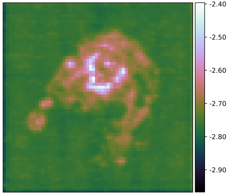


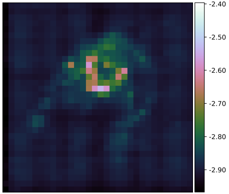
LCI$-$ $<$ LCI $>$





Posterior Standard Deviation



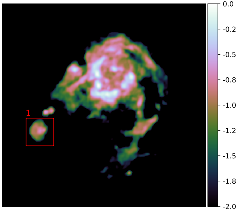


Pixel size 4×4 Pixel size 8×8

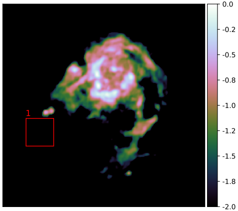
Computation time reduced by a factor of 10^3

Hypothesis test

Scalable hypothesis testing for structure in the reconstruction

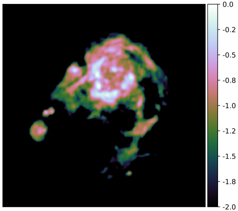


MAP reconstruction

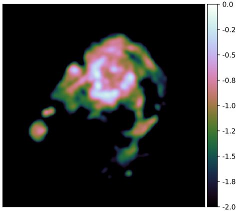


Inpainted surrogate

Is the blob physical? $\rightarrow$ Yes



MAP reconstruction



Blurred substructure

Is the substructure physical? $\rightarrow$ Yes

Conclusions

- **Scalable uncertainty quantification**
  - We exploit a concentration phenomenon of log-concave posteriors
  - Focus on hypothesis test and local credible intervals
- Only rely on **optimisation** to compute the MAP and avoid sampling
- We used **learned convex regularisers**
  - Considerably decreased reconstruction errors
  - Improved quality of the posterior St Dev

Conclusions

- **Scalable uncertainty quantification**
  - We exploit a concentration phenomenon of log-concave posteriors
  - Focus on hypothesis test and local credible intervals
- Only rely on **optimisation** to compute the MAP and avoid sampling
- We used **learned convex regularisers**
  - Considerably decreased reconstruction errors
  - Improved quality of the posterior St Dev

Conclusions

- **Scalable uncertainty quantification**
  - We exploit a concentration phenomenon of log-concave posteriors
  - Focus on hypothesis test and local credible intervals

- Only rely on **optimisation** to compute the MAP and avoid sampling

- We used **learned convex regularisers**
  - Considerably decreased reconstruction errors
  - Improved quality of the posterior St Dev

Conclusions

- **Scalable uncertainty quantification**
  - We exploit a concentration phenomenon of log-concave posteriors
  - Focus on hypothesis test and local credible intervals
- Only rely on **optimisation** to compute the MAP and avoid sampling
- We used **learned convex regularisers**
  - Considerably decreased reconstruction errors
  - Improved quality of the posterior St Dev

References

Cai, X., Pereyra, M., & McEwen, J. D. (2018). Uncertainty quantification for radio interferometric imaging – I. Proximal MCMC methods. Monthly Notices of the Royal Astronomical Society, 480(3), 4154–4169.

Goujon, A., Neumayer, S., Bohra, P., Ducotterd, S., & Unser, M. (2022). A Neural-Network-Based Convex Regularizer for Image Reconstruction. arXiv e-prints, Article arXiv:2211.12461, arXiv:2211.12461.

Pereyra, M. (2016). Proximal markov chain monte carlo algorithms. Statistics and Computing, 26(4), 745–760.

Pereyra, M. (2017). Maximum-a-posteriori estimation with bayesian confidence regions. SIAM Journal on Imaging Sciences, 10(1), 285–302.

Pereyra, M., Mieles, L. V., & Zygalakis, K. C. (2020). Accelerating proximal markov chain monte carlo by using an explicit stabilized method. SIAM Journal on Imaging Sciences, 13(2), 905–935.