High-dimensional uncertainty quantification

for radio interferometric imaging

Jason McEwen
www.jasonmcewen.org
@jasonmcewen

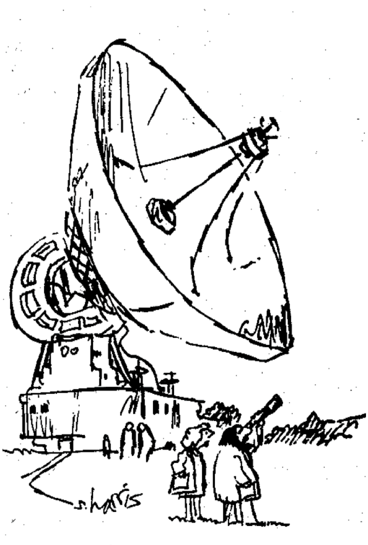
Mullard Space Science Laboratory (MSSL)
University College London (UCL)

with Xiaohao Cai (MSSL) and Marcelo Pereyra (HWU)

Cai, Pereyra & McEwen (2017a): arXiv:1711.04818
Cai, Pereyra & McEwen (2017b): arXiv:1711.04819

Workshop on Uncertainty Quantification and Computational Imaging,
International Centre for Mathematical Sciences (ICMS), Edinburgh
April 2018

Radio telescopes are big!



"Just checking."

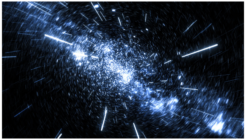
Radio telescopes are big!

Radio interferometric telescopes

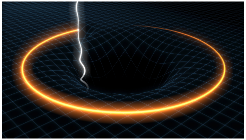
Very Large Array (VLA) in New Mexico

Next-generation of radio interferometry rapidly approaching

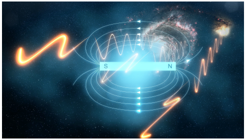
- Next-generation of radio interferometric telescopes will provide orders of magnitude improvement in sensitivity.
- Unlock broad range of science goals.



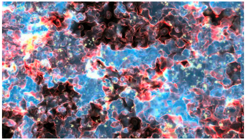
(a) **Dark energy**



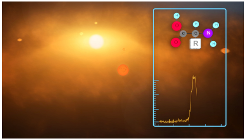
(b) **General relativity**



(c) **Cosmic magnetism**



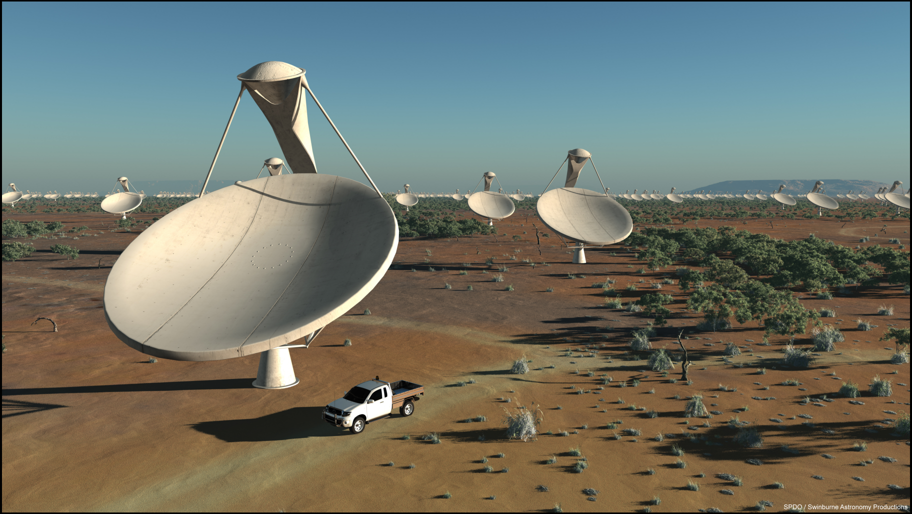
(d) **Epoch of reionization**



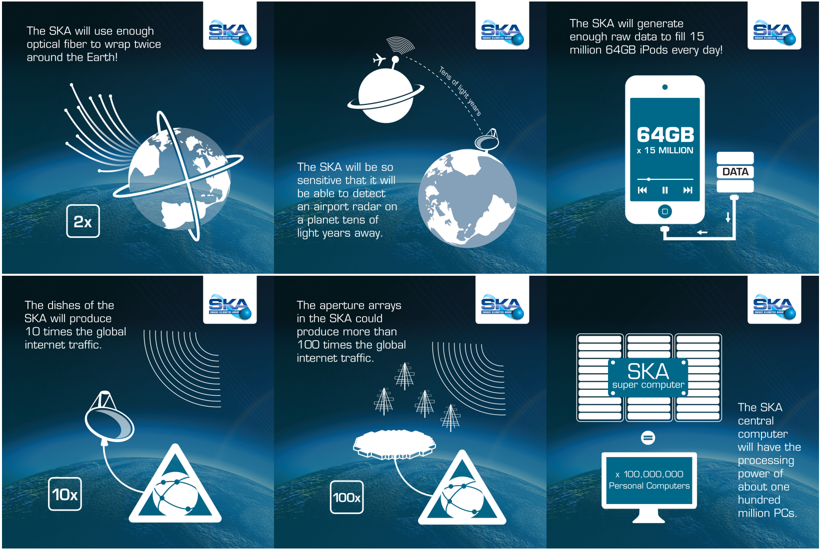
(e) **Exoplanets**

Figure: SKA science goals. **[Credit: SKA Organisation]**

Square Kilometre Array (SKA)



The SKA poses a considerable big-data challenge



The SKA poses a considerable big-data challenge

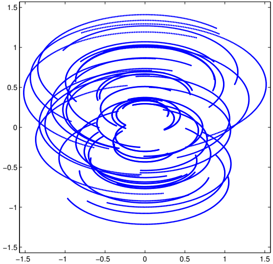
Outline

1. Radio interferometric imaging
2. Proximal MCMC sampling and uncertainty quantification
3. MAP estimation and uncertainty quantification

Outline

Radio interferometric telescopes acquire "Fourier" measurements

"Fourier"
Measurements



Radio interferometric inverse problem

- Consider the ill-posed inverse problem of radio interferometric imaging:

$$\boxed{\boldsymbol{y} = \boldsymbol{\Phi} \boldsymbol{x} + \boldsymbol{n}},$$

where $\boldsymbol{y}$ are the measured visibilities, $\boldsymbol{\Phi}$ is the linear measurement operator, $\boldsymbol{x}$ is the underlying image and $\boldsymbol{n}$ is instrumental noise.

- Measurement operator, *e.g.* $\boxed{\boldsymbol{\Phi} = \mathbf{GFA}}$, may incorporate:
  - primary beam $\mathbf{A}$ of the telescope;
  - Fourier transform $\mathbf{F}$;
  - convolutional de-gridding $\mathbf{G}$ to interpolate to continuous uv-coordinates;
  - direction-dependent effects (DDEs)...

Interferometric imaging: recover an image from noisy and incomplete Fourier measurements.

Radio interferometric inverse problem

- Consider the ill-posed inverse problem of radio interferometric imaging:

$$\boxed{\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{n}},$$

where $\boldsymbol{y}$ are the measured visibilities, $\boldsymbol{\Phi}$ is the linear measurement operator, $\boldsymbol{x}$ is the underlying image and $\boldsymbol{n}$ is instrumental noise.

- Measurement operator, *e.g.* $\boxed{\boldsymbol{\Phi} = \mathbf{GFA}}$, may incorporate:
  - primary beam $\mathbf{A}$ of the telescope;
  - Fourier transform $\mathbf{F}$;
  - convolutional de-gridding $\mathbf{G}$ to interpolate to continuous uv-coordinates;
  - direction-dependent effects (DDEs)...

Interferometric imaging: recover an image from noisy and incomplete Fourier measurements.

Radio interferometric inverse problem

- Consider the ill-posed inverse problem of radio interferometric imaging:

$$\boxed{\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{n}},$$

where $\boldsymbol{y}$ are the measured visibilities, $\boldsymbol{\Phi}$ is the linear measurement operator, $\boldsymbol{x}$ is the underlying image and $\boldsymbol{n}$ is instrumental noise.

- Measurement operator, *e.g.* $\boxed{\boldsymbol{\Phi} = \mathbf{GFA}}$, may incorporate:
  - primary beam **A** of the telescope;
  - Fourier transform **F**;
  - convolutional de-gridding **G** to interpolate to continuous uv-coordinates;
  - direction-dependent effects (DDEs)...

Interferometric imaging: recover an image from noisy and incomplete Fourier measurements.

Sparse regularisation

Synthesis and analysis frameworks

- Sparse synthesis regularisation problem:

$$\boxed{\boldsymbol{x}_{\text{synthesis}} = \boldsymbol{\Psi} \times \arg\min_{\boldsymbol{\alpha}} \left[\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \right]}$$

Synthesis framework

where consider sparsifying (*e.g.* wavelet) representation of image: $\boxed{\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}}$.

- Typically sparsity assumption justified by analysing example signals in transformed domain.
- Different to synthesising signals.
- Suggests sparse analysis regularisation problem (Elad *et al.* 2007, Nam *et al.* 2012):

$$\boxed{\boldsymbol{x}_{\text{analysis}} = \arg\min_{\boldsymbol{x}} \left[\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{\Psi}^\dagger \boldsymbol{x}\|_1 \right]}$$

Analysis framework

(For orthogonal bases the two approaches are identical but otherwise very different.)

Sparse regularisation

Synthesis and analysis frameworks

- Sparse synthesis regularisation problem:

$$\boxed{\boldsymbol{x}_{\text{synthesis}} = \boldsymbol{\Psi} \times \arg\min_{\boldsymbol{\alpha}} \left[\left\| \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha} \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha} \right\|_1 \right]}$$

Synthesis framework

where consider sparsifying (*e.g.* wavelet) representation of image: $\boxed{\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}}$.

- Typically sparsity assumption justified by analysing example signals in transformed domain.
- Different to synthesising signals.
- Suggests sparse analysis regularisation problem (Elad *et al.* 2007, Nam *et al.* 2012):

$$\boxed{\boldsymbol{x}_{\text{analysis}} = \arg\min_{\boldsymbol{x}} \left[\left\| \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x} \right\|_2^2 + \lambda \left\| \boldsymbol{\Psi}^{\dagger}\boldsymbol{x} \right\|_1 \right]}$$

Analysis framework

(For orthogonal bases the two approaches are identical but otherwise very different.)

Sparse regularisation

Synthesis and analysis frameworks

- Sparse synthesis regularisation problem:

$$\boxed{\boldsymbol{x}_{\text{synthesis}} = \boldsymbol{\Psi} \times \underset{\boldsymbol{\alpha}}{\arg\min} \left[\left\| \boldsymbol{y} - \boldsymbol{\Phi\Psi\alpha} \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha} \right\|_1 \right]}$$

Synthesis framework

where consider sparsifying (*e.g.* wavelet) representation of image: $\boxed{\boldsymbol{x} = \boldsymbol{\Psi\alpha}}$.

- Typically sparsity assumption justified by analysing example signals in transformed domain.
- Different to synthesising signals.
- Suggests sparse analysis regularisation problem (Elad *et al.* 2007, Nam *et al.* 2012):

$$\boxed{\boldsymbol{x}_{\text{analysis}} = \underset{\boldsymbol{x}}{\arg\min} \left[\left\| \boldsymbol{y} - \boldsymbol{\Phi x} \right\|_2^2 + \lambda \left\| \boldsymbol{\Psi}^\dagger \boldsymbol{x} \right\|_1 \right]}$$

Analysis framework

(For orthogonal bases the two approaches are identical but otherwise very different.)

Sparse regularisation

SARA algorithm

- Sparsity averaging reweighted analysis (SARA)
  (Carrillo, McEwen & Wiaux 2012; Carrillo, McEwen, Van De Ville, Thiran & Wiaux 2013).

- Overcomplete dictionary composed of a concatenation of orthonormal bases:

$$\boxed{\boldsymbol{\Psi} = [\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \ldots, \boldsymbol{\Psi}_q]}$$

  with following bases: Dirac (*i.e.* pixel basis); Haar wavelets (promotes gradient sparsity); Daubechies wavelets two to eight ⇒ concatenation of 9 bases.

- Promote average sparsity by solving the constrained reweighted ℓ_1 analysis problem:

$$\boxed{\min_{\boldsymbol{x} \in \mathbb{R}^N} \|\mathbf{W}\boldsymbol{\Psi}^\dagger \boldsymbol{x}\|_1 \quad \text{subject to} \quad \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2 \leq \epsilon \quad \text{and} \quad \boldsymbol{x} \geq 0} \quad \text{SARA}$$

Sparse regularisation
SARA algorithm

- Sparsity averaging reweighted analysis (SARA)
  (Carrillo, McEwen & Wiaux 2012; Carrillo, McEwen, Van De Ville, Thiran & Wiaux 2013).

- Overcomplete dictionary composed of a concatenation of orthonormal bases:

$$\boxed{\boldsymbol{\Psi} = \left[\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \ldots, \boldsymbol{\Psi}_q\right]}$$

  with following bases: Dirac (*i.e.* pixel basis); Haar wavelets (promotes gradient sparsity); Daubechies wavelets two to eight $\Rightarrow$ concatenation of 9 bases.

- Promote average sparsity by solving the constrained reweighted ℓ_1 analysis problem:

$$\boxed{\min_{\boldsymbol{x} \in \mathbb{R}^N} \|\mathbf{W}\boldsymbol{\Psi}^\dagger \boldsymbol{x}\|_1 \quad \text{subject to} \quad \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2 \leq \epsilon \quad \text{and} \quad \boldsymbol{x} \geq 0}$$ SARA

Sparse regularisation
SARA algorithm

- Sparsity averaging reweighted analysis (SARA)
  (Carrillo, McEwen & Wiaux 2012; Carrillo, McEwen, Van De Ville, Thiran & Wiaux 2013).

- Overcomplete dictionary composed of a concatenation of orthonormal bases:

$$\boxed{\boldsymbol{\Psi} = \left[\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \ldots, \boldsymbol{\Psi}_q\right]}$$

  with following bases: Dirac (*i.e.* pixel basis); Haar wavelets (promotes gradient sparsity); Daubechies wavelets two to eight $\Rightarrow$ concatenation of 9 bases.

- Promote average sparsity by solving the constrained reweighted ℓ_1 analysis problem:

$$\boxed{\min_{\boldsymbol{x}\in\mathbb{R}^N} \|\mathbf{W}\boldsymbol{\Psi}^\dagger\boldsymbol{x}\|_1 \quad \text{subject to} \quad \|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\|_2 \leq \epsilon \quad \text{and} \quad \boldsymbol{x} \geq 0} \quad \text{SARA}$$

Public open-source codes

PURIFY code

http://basp-group.github.io/purify/

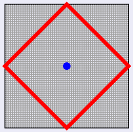
Next-generation radio interferometric imaging

Carrillo, McEwen, Wiaux, Pratley, d'Avezac

PURIFY is an open-source code that provides functionality to perform radio interferometric imaging, leveraging recent developments in the field of compressive sensing and convex optimisation.

SOPT code

http://basp-group.github.io/sopt/



Sparse OPTimisation

Carrillo, McEwen, Wiaux, Kartik, d'Avezac, Pratley, Perez-Suarez

SOPT is an open-source code that provides functionality to perform sparse optimisation using state-of-the-art convex optimisation algorithms.

Imaging observations from the VLA and ATCA with PURIFY

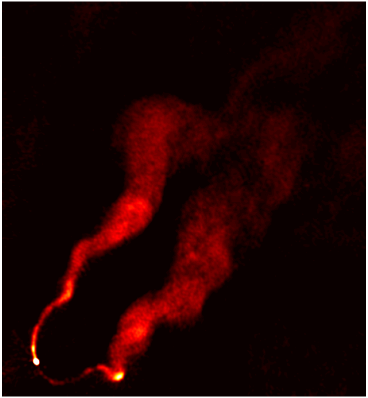
(a) NRAO Very Large Array (VLA)

(b) Australia Telescope Compact Array (ATCA)

Figure: Radio interferometric telescopes considered

PURIFY reconstruction

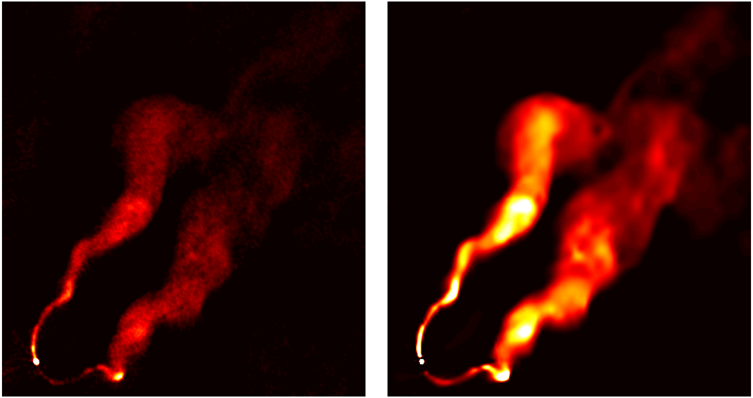
VLA observation of 3C129



(a) CLEAN (uniform)

Figure: 3C129 recovered images (Pratley, McEwen, *et al.* 2016)

PURIFY reconstruction
VLA observation of 3C129

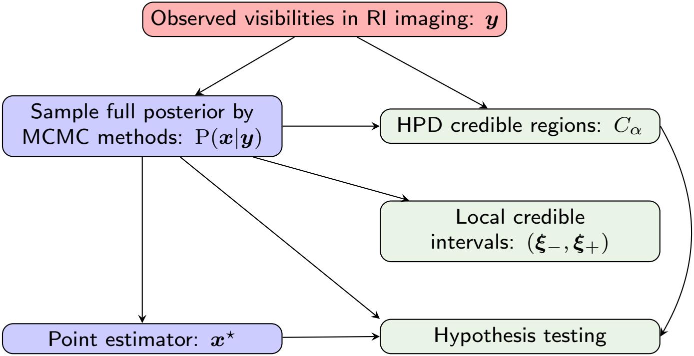


(a) CLEAN (uniform) (b) PURIFY

Figure: 3C129 recovered images (Pratley, McEwen, *et al.* 2016)

Outline

MCMC sampling and uncertainty quantification



MCMC sampling the full posterior distribution

- Sample full posterior distribution $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y})$.

- MCMC methods for high-dimensional problems (like interferometric imaging):
    - Gibbs sampling (sample from conditional distributions)
    - Hamiltonian MC (HMC) sampling (exploit gradients)
    - Metropolis adjusted Langevin algorithm (MALA) sampling (exploit gradients)

Require MCMC approach to support sparsity-promoting priors, which shown to be highly effective.

MCMC sampling the full posterior distribution

- Sample full posterior distribution $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y})$.
- MCMC methods for high-dimensional problems (like interferometric imaging):
  - Gibbs sampling (sample from conditional distributions)
  - Hamiltonian MC (HMC) sampling (exploit gradients)
  - Metropolis adjusted Langevin algorithm (MALA) sampling (exploit gradients)

Require MCMC approach to support sparsity-promoting priors, which shown to be highly effective.

MCMC sampling the full posterior distribution

- Sample full posterior distribution $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y})$.
- MCMC methods for high-dimensional problems (like interferometric imaging):
  - Gibbs sampling (sample from conditional distributions)
  - Hamiltonian MC (HMC) sampling (exploit gradients)
  - Metropolis adjusted Langevin algorithm (MALA) sampling (exploit gradients)

Require MCMC approach to support sparsity-promoting priors, which shown to be highly effective.

MCMC sampling with gradients
Langevin dynamics

- Consider posteriors of the following form:

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \underbrace{\boxed{\pi(\boldsymbol{x})}}_{\text{Posterior}} \propto \exp\big(-\underbrace{\boxed{g(\boldsymbol{x})}}_{\text{Smooth}}\big)$$

- If $g(x)$ differentiable can adopt MALA (Langevin dynamics).
- Based on Langevin diffusion process $\mathcal{L}(t)$, with π as stationary distribution:

$$\mathrm{d}\mathcal{L}(t) = \frac{1}{2}\nabla \log \pi\big(\mathcal{L}(t)\big)\mathrm{d}t + \mathrm{d}\mathcal{W}(t), \quad \mathcal{L}(0) = l_0$$

  where $\mathcal{W}$ is Brownian motion.
- Need gradients so cannot support sparsity-promoting priors.

MCMC sampling with gradients

Langevin dynamics

- Consider posteriors of the following form:

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \underbrace{\boxed{\pi(\boldsymbol{x})}}_{\text{Posterior}} \propto \exp\big(-\underbrace{\boxed{g(\boldsymbol{x})}}_{\text{Smooth}}\big)$$

- If $g(\boldsymbol{x})$ differentiable can adopt MALA (Langevin dynamics).
- Based on Langevin diffusion process $\mathcal{L}(t)$, with π as stationary distribution:

$$\mathrm{d}\mathcal{L}(t) = \frac{1}{2}\nabla \log \pi\big(\mathcal{L}(t)\big)\mathrm{d}t + \mathrm{d}\mathcal{W}(t), \quad \mathcal{L}(0) = l_0$$

where $\mathcal{W}$ is Brownian motion.

- Need gradients so cannot support sparsity-promoting priors.

MCMC sampling with gradients

Langevin dynamics

- Consider posteriors of the following form:

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \underbrace{\boxed{\pi(\boldsymbol{x})}}_{\text{Posterior}} \propto \exp\big(-\underbrace{\boxed{g(\boldsymbol{x})}}_{\text{Smooth}}\big)$$

- If $g(\boldsymbol{x})$ differentiable can adopt MALA (Langevin dynamics).
- Based on Langevin diffusion process $\mathcal{L}(t)$, with π as stationary distribution:

$$\mathrm{d}\mathcal{L}(t) = \frac{1}{2}\nabla \log \pi\big(\mathcal{L}(t)\big)\mathrm{d}t + \mathrm{d}\mathcal{W}(t), \quad \mathcal{L}(0) = l_0$$

  where $\mathcal{W}$ is Brownian motion.
- Need gradients so cannot support sparsity-promoting priors.

MCMC sampling with gradients

Langevin dynamics

- Consider posteriors of the following form:

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \underbrace{\boxed{\pi(\boldsymbol{x})}}_{\text{Posterior}} \propto \exp\big(-\underbrace{\boxed{g(\boldsymbol{x})}}_{\text{Smooth}}\big)$$

- If $g(\boldsymbol{x})$ differentiable can adopt MALA (Langevin dynamics).
- Based on Langevin diffusion process $\mathcal{L}(t)$, with π as stationary distribution:

$$\mathrm{d}\mathcal{L}(t) = \frac{1}{2} \underbrace{\boxed{\nabla \log \pi\big(\mathcal{L}(t)\big)}}_{\text{Gradient}} \mathrm{d}t + \mathrm{d}\mathcal{W}(t), \quad \mathcal{L}(0) = l_0$$

  where $\mathcal{W}$ is Brownian motion.

- Need gradients so cannot support sparsity-promoting priors.

Proximal MALA

Moreau approximation

- Moreau approximation of $f(\boldsymbol{x}) \propto \exp(-g(\boldsymbol{x}))$:

$$\boxed{f_\lambda^{\mathsf{MA}}(\boldsymbol{x}) = \sup_{\boldsymbol{u}\in\mathbb{R}^N} f(\boldsymbol{u}) \exp\left(-\frac{\|\boldsymbol{u}-\boldsymbol{x}\|^2}{2\lambda}\right)}$$

- Important properties of $f_\lambda^{\mathsf{MA}}(\boldsymbol{x})$:

  1. As $\lambda \to 0$, $f_\lambda^{\mathsf{MA}}(\boldsymbol{x}) \to f(\boldsymbol{x})$
  2. $\nabla \log f_\lambda^{\mathsf{MA}}(\boldsymbol{x}) = (\mathrm{prox}_g^\lambda(\boldsymbol{x}) - \boldsymbol{x})/\lambda$

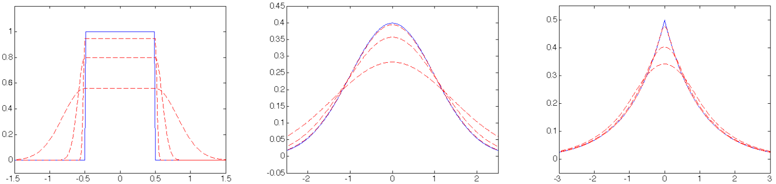


Figure: Illustration of Moreau approximations [Credit: Pereyra 2016a]

Proximal MALA

Moreau approximation

- Moreau approximation of $f(\boldsymbol{x}) \propto \exp(-g(\boldsymbol{x}))$:

$$\boxed{f_\lambda^{\mathsf{MA}}(\boldsymbol{x}) = \sup_{\boldsymbol{u} \in \mathbb{R}^N} f(\boldsymbol{u}) \exp\left(-\frac{\|\boldsymbol{u}-\boldsymbol{x}\|^2}{2\lambda}\right)}$$

- Important properties of $f_\lambda^{\mathsf{MA}}(\boldsymbol{x})$:

  1. As $\lambda \rightarrow 0$, $f_\lambda^{\mathsf{MA}}(\boldsymbol{x}) \rightarrow f(\boldsymbol{x})$
  2. $\nabla \log f_\lambda^{\mathsf{MA}}(\boldsymbol{x}) = (\text{prox}_g^\lambda(\boldsymbol{x}) - \boldsymbol{x})/\lambda$

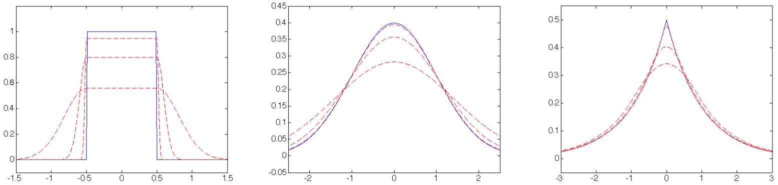


Figure: Illustration of Moreau approximations [Credit: Pereyra 2016a]

Proximal MALA

MCMC sampling

Proximal Metropolis adjusted Langevin algorithm (Px-MALA)

Pereyra (2016a)

- Consider log-convex posteriors: $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \pi(\boldsymbol{x}) \propto \exp\big(-\underbrace{\boxed{g(\boldsymbol{x})}}_{\text{Convex}}\big)$.
- Langevin diffusion process $\mathcal{L}(t)$, with π as stationary distribution ($\mathcal{W}$ Brownian motion):

$$\mathrm{d}\mathcal{L}(t) = \frac{1}{2}\nabla \log \pi\big(\mathcal{L}(t)\big)\mathrm{d}t + \mathrm{d}\mathcal{W}(t), \quad \mathcal{L}(0) = l_0 \,.$$

- Euler discretisation and apply Moreau approximation to π:

$$\boldsymbol{l}^{(m+1)} = \boldsymbol{l}^{(m)} + \frac{\delta}{2} \boxed{\nabla \log \pi(\boldsymbol{l}^{(m)})} + \sqrt{\delta}\boldsymbol{w}^{(m)} \,.$$

$$\nabla \log \pi_\lambda(\boldsymbol{x}) = (\mathrm{prox}_g^\lambda(\boldsymbol{x}) - \boldsymbol{x})/\lambda$$

- Metropolis-Hastings accept-reject step.

Proximal MALA

MCMC sampling

Proximal Metropolis adjusted Langevin algorithm (Px-MALA)

Pereyra (2016a)

- Consider log-convex posteriors: $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \pi(\boldsymbol{x}) \propto \exp\big(-\underbrace{\boxed{g(\boldsymbol{x})}}_{\text{Convex}}\big)$.
- Langevin diffusion process $\mathcal{L}(t)$, with π as stationary distribution ($\mathcal{W}$ Brownian motion):

$$\mathrm{d}\mathcal{L}(t) = \frac{1}{2}\nabla \log \pi\big(\mathcal{L}(t)\big)\mathrm{d}t + \mathrm{d}\mathcal{W}(t), \quad \mathcal{L}(0) = l_0 \ .$$

- Euler discretisation and apply Moreau approximation to π:

$$\boldsymbol{l}^{(m+1)} = \boldsymbol{l}^{(m)} + \frac{\delta}{2}\boxed{\nabla \log \pi(\boldsymbol{l}^{(m)})} + \sqrt{\delta}\boldsymbol{w}^{(m)} \ .$$

$$\nabla \log \pi_\lambda(\boldsymbol{x}) = (\mathrm{prox}_g^\lambda(\boldsymbol{x}) - \boldsymbol{x})/\lambda$$

- Metropolis-Hastings accept-reject step.

Proximal MALA

MCMC sampling

Proximal Metropolis adjusted Langevin algorithm (Px-MALA)

Pereyra (2016a)

- Consider log-convex posteriors: $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \pi(\boldsymbol{x}) \propto \exp\big(-\underbrace{\boxed{g(\boldsymbol{x})}}_{\text{Convex}}\big)$.
- Langevin diffusion process $\mathcal{L}(t)$, with π as stationary distribution ($\mathcal{W}$ Brownian motion):

$$\mathrm{d}\mathcal{L}(t) = \frac{1}{2}\nabla \log \pi\big(\mathcal{L}(t)\big)\mathrm{d}t + \mathrm{d}\mathcal{W}(t), \quad \mathcal{L}(0) = l_0 \, .$$

- Euler discretisation and apply Moreau approximation to π:

$$\boldsymbol{l}^{(m+1)} = \boldsymbol{l}^{(m)} + \frac{\delta}{2} \boxed{\nabla \log \pi(\boldsymbol{l}^{(m)})} + \sqrt{\delta}\boldsymbol{w}^{(m)} \, .$$

$$\nabla \log \pi_\lambda(\boldsymbol{x}) = (\mathrm{prox}_g^\lambda(\boldsymbol{x}) - \boldsymbol{x})/\lambda$$

- Metropolis-Hastings accept-reject step.

Proximal MALA

MCMC sampling

Proximal Metropolis adjusted Langevin algorithm (Px-MALA)

Pereyra (2016a)

- Consider log-convex posteriors: $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \pi(\boldsymbol{x}) \propto \exp\big(-\underbrace{\boxed{g(\boldsymbol{x})}}_{\text{Convex}}\big)$.
- Langevin diffusion process $\mathcal{L}(t)$, with π as stationary distribution ($\mathcal{W}$ Brownian motion):

$$\mathrm{d}\mathcal{L}(t) = \frac{1}{2}\nabla \log \pi\big(\mathcal{L}(t)\big)\mathrm{d}t + \mathrm{d}\mathcal{W}(t), \quad \mathcal{L}(0) = l_0 \, .$$

- Euler discretisation and apply Moreau approximation to π:

$$\boldsymbol{l}^{(m+1)} = \boldsymbol{l}^{(m)} + \frac{\delta}{2}\boxed{\nabla \log \pi(\boldsymbol{l}^{(m)})} + \sqrt{\delta}\boldsymbol{w}^{(m)} \, .$$

$$\nabla \log \pi_\lambda(\boldsymbol{x}) = (\mathrm{prox}_g^\lambda(\boldsymbol{x}) - \boldsymbol{x})/\lambda$$

- Metropolis-Hastings accept-reject step.

Proximal MALA

Computing proximity operators for the analysis case

- Recall posterior: $\pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$.
- Let $\bar{g}(\boldsymbol{x}) = \bar{f}_1(\boldsymbol{x}) + \bar{f}_2(\boldsymbol{x})$, where $\boxed{\bar{f}_1(\boldsymbol{x}) = \mu\|\boldsymbol{\Psi}^\dagger \boldsymbol{x}\|_1}$ and $\boxed{\bar{f}_2(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2/2\sigma^2}$.

Prior Likelihood

- Must solve an optimisation problem for each iteration!

$$\boxed{\mathrm{prox}_{\bar{g}}^{\delta/2}(\boldsymbol{x}) = \operatorname*{argmin}_{\boldsymbol{u}\in\mathbb{R}^N} \left\{ \mu\|\boldsymbol{\Psi}^\dagger \boldsymbol{u}\|_1 + \frac{\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2^2}{2\sigma^2} + \frac{\|\boldsymbol{u} - \boldsymbol{x}\|_2^2}{\delta} \right\}}\,.$$

- Taylor expansion at point $\boldsymbol{x}$: $\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2^2 \approx \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + 2(\boldsymbol{u} - \boldsymbol{x})^\top \boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})$.
- Then proximity operator approximated by

$$\boxed{\mathrm{prox}_{\bar{g}}^{\delta/2}(\boldsymbol{x}) \approx \mathrm{prox}_{\bar{f}_1}^{\delta/2}\left(\boldsymbol{x} - \delta\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})/2\sigma^2\right)}\,.$$

Single forward-backward iteration

- Analytic approximation:

$$\boxed{\mathrm{prox}_{\bar{g}}^{\delta/2}(\boldsymbol{x}) \approx \bar{\boldsymbol{v}} + \boldsymbol{\Psi}\left(\mathrm{soft}_{\mu\delta/2}(\boldsymbol{\Psi}^\dagger \bar{\boldsymbol{v}}) - \boldsymbol{\Psi}^\dagger \bar{\boldsymbol{v}}\right)}\,, \quad \text{where } \bar{\boldsymbol{v}} = \boldsymbol{x} - \delta\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})/2\sigma^2.$$

Proximal MALA

Computing proximity operators for the analysis case

- Recall posterior: $\pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$.
- Let $\bar{g}(\boldsymbol{x}) = \bar{f}_1(\boldsymbol{x}) + \bar{f}_2(\boldsymbol{x})$, where $\boxed{\bar{f}_1(\boldsymbol{x}) = \mu \|\boldsymbol{\Psi}^\dagger \boldsymbol{x}\|_1}$ and $\boxed{\bar{f}_2(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 / 2\sigma^2}$.

  Prior Likelihood

- Must solve an optimisation problem for each iteration!

$$\boxed{\mathrm{prox}_{\bar{g}}^{\delta/2}(\boldsymbol{x}) = \underset{\boldsymbol{u} \in \mathbb{R}^N}{\mathrm{argmin}} \left\{ \mu \|\boldsymbol{\Psi}^\dagger \boldsymbol{u}\|_1 + \frac{\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2^2}{2\sigma^2} + \frac{\|\boldsymbol{u} - \boldsymbol{x}\|_2^2}{\delta} \right\}}.$$

- Taylor expansion at point $\boldsymbol{x}$: $\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2^2 \approx \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + 2(\boldsymbol{u} - \boldsymbol{x})^\top \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})$.
- Then proximity operator approximated by

$$\boxed{\mathrm{prox}_{\bar{g}}^{\delta/2}(\boldsymbol{x}) \approx \mathrm{prox}_{\bar{f}_1}^{\delta/2}\left(\boldsymbol{x} - \delta \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})/2\sigma^2\right)}.$$

Single forward-backward iteration

- Analytic approximation:

$$\boxed{\mathrm{prox}_{\bar{g}}^{\delta/2}(\boldsymbol{x}) \approx \bar{\boldsymbol{v}} + \boldsymbol{\Psi}\left(\mathrm{soft}_{\mu\delta/2}(\boldsymbol{\Psi}^\dagger \bar{\boldsymbol{v}}) - \boldsymbol{\Psi}^\dagger \bar{\boldsymbol{v}}\right)},$$ where $\bar{\boldsymbol{v}} = \boldsymbol{x} - \delta \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})/2\sigma^2$.

Proximal MALA

Computing proximity operators for the analysis case

- Recall posterior: $\pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$.
- Let $\bar{g}(\boldsymbol{x}) = \bar{f}_1(\boldsymbol{x}) + \bar{f}_2(\boldsymbol{x})$, where $\boxed{\bar{f}_1(\boldsymbol{x}) = \mu \|\boldsymbol{\Psi}^\dagger \boldsymbol{x}\|_1}$ and $\boxed{\bar{f}_2(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 / 2\sigma^2}$.

  Prior Likelihood

- Must solve an optimisation problem for each iteration!

$$\boxed{\operatorname{prox}_{\bar{g}}^{\delta/2}(\boldsymbol{x}) = \underset{\boldsymbol{u} \in \mathbb{R}^N}{\operatorname{argmin}} \left\{ \mu \|\boldsymbol{\Psi}^\dagger \boldsymbol{u}\|_1 + \frac{\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2^2}{2\sigma^2} + \frac{\|\boldsymbol{u} - \boldsymbol{x}\|_2^2}{\delta} \right\}}.$$

- Taylor expansion at point $\boldsymbol{x}$: $\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2^2 \approx \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + 2(\boldsymbol{u} - \boldsymbol{x})^\top \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})$.
- Then proximity operator approximated by

$$\boxed{\operatorname{prox}_{\bar{g}}^{\delta/2}(\boldsymbol{x}) \approx \operatorname{prox}_{\bar{f}_1}^{\delta/2}\left(\boldsymbol{x} - \delta \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})/2\sigma^2\right)}.$$

Single forward-backward iteration

- Analytic approximation:

$$\boxed{\operatorname{prox}_{\bar{g}}^{\delta/2}(\boldsymbol{x}) \approx \bar{\boldsymbol{v}} + \boldsymbol{\Psi}\left(\operatorname{soft}_{\mu\delta/2}(\boldsymbol{\Psi}^\dagger \bar{\boldsymbol{v}}) - \boldsymbol{\Psi}^\dagger \bar{\boldsymbol{v}}\right)}, \quad \text{where } \bar{\boldsymbol{v}} = \boldsymbol{x} - \delta \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})/2\sigma^2.$$

Proximal MALA

Computing proximity operators for the analysis case

- Recall posterior: $\pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$.
- Let $\bar{g}(\boldsymbol{x}) = \bar{f}_1(\boldsymbol{x}) + \bar{f}_2(\boldsymbol{x})$, where $\boxed{\bar{f}_1(\boldsymbol{x}) = \mu \|\boldsymbol{\Psi}^\dagger \boldsymbol{x}\|_1}$ and $\boxed{\bar{f}_2(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 / 2\sigma^2}$.

  Prior Likelihood

- Must solve an optimisation problem for each iteration!

$$\boxed{\mathrm{prox}_{\bar{g}}^{\delta/2}(\boldsymbol{x}) = \operatorname*{argmin}_{\boldsymbol{u} \in \mathbb{R}^N} \left\{ \mu \|\boldsymbol{\Psi}^\dagger \boldsymbol{u}\|_1 + \frac{\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2^2}{2\sigma^2} + \frac{\|\boldsymbol{u} - \boldsymbol{x}\|_2^2}{\delta} \right\}} .$$

- Taylor expansion at point $\boldsymbol{x}$: $\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2^2 \approx \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + 2(\boldsymbol{u} - \boldsymbol{x})^\top \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})$.
- Then proximity operator approximated by

$$\boxed{\mathrm{prox}_{\bar{g}}^{\delta/2}(\boldsymbol{x}) \approx \mathrm{prox}_{\bar{f}_1}^{\delta/2} \left(\boldsymbol{x} - \delta \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y}) / 2\sigma^2 \right)} .$$

Single forward-backward iteration

- Analytic approximation:

$$\boxed{\mathrm{prox}_{\bar{g}}^{\delta/2}(\boldsymbol{x}) \approx \bar{\boldsymbol{v}} + \boldsymbol{\Psi} \left(\mathrm{soft}_{\mu\delta/2}(\boldsymbol{\Psi}^\dagger \bar{\boldsymbol{v}}) - \boldsymbol{\Psi}^\dagger \bar{\boldsymbol{v}} \right)} , \quad \text{where } \bar{\boldsymbol{v}} = \boldsymbol{x} - \delta \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y}) / 2\sigma^2 .$$

Proximal MALA

Computing proximity operators for the synthesis case

- Recall posterior: $\pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$.
- Let $\hat{g}(\boldsymbol{x}(\boldsymbol{a})) = \hat{f}_1(\boldsymbol{a}) + \hat{f}_2(\boldsymbol{a})$, where $\boxed{\hat{f}_1(\boldsymbol{a}) = \mu\|\boldsymbol{a}\|_1}$ and $\boxed{\hat{f}_2(\boldsymbol{a}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2/2\sigma^2}$.

  Prior Likelihood

- Must solve an optimisation problem for each iteration!

$$\boxed{\operatorname{prox}_{\hat{g}}^{\delta/2}(\boldsymbol{a}) = \underset{\boldsymbol{u}\in\mathbb{R}^L}{\operatorname{argmin}} \left\{ \mu\|\boldsymbol{u}\|_1 + \frac{\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{u}\|_2^2}{2\sigma^2} + \frac{\|\boldsymbol{u} - \boldsymbol{a}\|_2^2}{\delta} \right\}} .$$

- Taylor expansion at point $\boldsymbol{a}$: $\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{u}\|_2^2 \approx \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2 + 2(\boldsymbol{u} - \boldsymbol{a})^\top \boldsymbol{\Psi}^\dagger \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a} - \boldsymbol{y})$.
- Then proximity operator approximated by

$$\boxed{\operatorname{prox}_{\hat{g}}^{\delta/2}(\boldsymbol{a}) \approx \operatorname{prox}_{\hat{f}_1}^{\delta/2}\left(\boldsymbol{a} - \delta\boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a} - \boldsymbol{y})/2\sigma^2\right)} .$$

Single forward-backward iteration

- Analytic approximation:

$$\boxed{\operatorname{prox}_{\hat{g}}^{\delta/2}(\boldsymbol{a}) \approx \operatorname{soft}_{\mu\delta/2}\left(\boldsymbol{a} - \delta\boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a} - \boldsymbol{y})/2\sigma^2\right)} .$$

Proximal MALA

Computing proximity operators for the synthesis case

- Recall posterior: $\pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$.
- Let $\hat{g}(\boldsymbol{x}(\boldsymbol{a})) = \hat{f}_1(\boldsymbol{a}) + \hat{f}_2(\boldsymbol{a})$, where $\underbrace{\boxed{\hat{f}_1(\boldsymbol{a}) = \mu \|\boldsymbol{a}\|_1}}_{\text{Prior}}$ and $\underbrace{\boxed{\hat{f}_2(\boldsymbol{a}) = \|\boldsymbol{y} - \boldsymbol{\Phi\Psi a}\|_2^2 / 2\sigma^2}}_{\text{Likelihood}}$.
- Must solve an optimisation problem for each iteration!

$$\boxed{\mathrm{prox}_{\hat{g}}^{\delta/2}(\boldsymbol{a}) = \underset{\boldsymbol{u} \in \mathbb{R}^L}{\mathrm{argmin}} \left\{ \mu \|\boldsymbol{u}\|_1 + \frac{\|\boldsymbol{y} - \boldsymbol{\Phi\Psi u}\|_2^2}{2\sigma^2} + \frac{\|\boldsymbol{u} - \boldsymbol{a}\|_2^2}{\delta} \right\}} .$$

- Taylor expansion at point $\boldsymbol{a}$: $\|\boldsymbol{y} - \boldsymbol{\Phi\Psi u}\|_2^2 \approx \|\boldsymbol{y} - \boldsymbol{\Phi\Psi a}\|_2^2 + 2(\boldsymbol{u} - \boldsymbol{a})^\top \boldsymbol{\Psi}^\dagger \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi\Psi a} - \boldsymbol{y})$.
- Then proximity operator approximated by

$$\underbrace{\boxed{\mathrm{prox}_{\hat{g}}^{\delta/2}(\boldsymbol{a}) \approx \mathrm{prox}_{\hat{f}_1}^{\delta/2}\left(\boldsymbol{a} - \delta \boldsymbol{\Psi}^\dagger \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi\Psi a} - \boldsymbol{y}) / 2\sigma^2\right)}}_{\text{Single forward-backward iteration}} .$$

- Analytic approximation:

$$\boxed{\mathrm{prox}_{\hat{g}}^{\delta/2}(\boldsymbol{a}) \approx \mathrm{soft}_{\mu\delta/2}\left(\boldsymbol{a} - \delta \boldsymbol{\Psi}^\dagger \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi\Psi a} - \boldsymbol{y}) / 2\sigma^2\right)} .$$

Proximal MALA

Computing proximity operators for the synthesis case

- Recall posterior: $\pi(\boldsymbol{x}) \propto \exp\bigl(-g(\boldsymbol{x})\bigr)$.
- Let $\hat{g}(\boldsymbol{x}(\boldsymbol{a})) = \hat{f}_1(\boldsymbol{a}) + \hat{f}_2(\boldsymbol{a})$, where $\boxed{\hat{f}_1(\boldsymbol{a}) = \mu \|\boldsymbol{a}\|_1}$ and $\boxed{\hat{f}_2(\boldsymbol{a}) = \|\boldsymbol{y} - \boldsymbol{\Phi\Psi a}\|_2^2 / 2\sigma^2}$.

  Prior Likelihood
- Must solve an optimisation problem for each iteration!

$$\boxed{\operatorname{prox}_{\hat{g}}^{\delta/2}(\boldsymbol{a}) = \underset{\boldsymbol{u} \in \mathbb{R}^L}{\operatorname{argmin}} \left\{ \mu \|\boldsymbol{u}\|_1 + \frac{\|\boldsymbol{y} - \boldsymbol{\Phi\Psi u}\|_2^2}{2\sigma^2} + \frac{\|\boldsymbol{u} - \boldsymbol{a}\|_2^2}{\delta} \right\}}\ .$$

- Taylor expansion at point $\boldsymbol{a}$: $\|\boldsymbol{y} - \boldsymbol{\Phi\Psi u}\|_2^2 \approx \|\boldsymbol{y} - \boldsymbol{\Phi\Psi a}\|_2^2 + 2(\boldsymbol{u} - \boldsymbol{a})^\top \boldsymbol{\Psi}^\dagger \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi\Psi a} - \boldsymbol{y})$.
- Then proximity operator approximated by

$$\boxed{\operatorname{prox}_{\hat{g}}^{\delta/2}(\boldsymbol{a}) \approx \operatorname{prox}_{\hat{f}_1}^{\delta/2} \left(\boldsymbol{a} - \delta \boldsymbol{\Psi}^\dagger \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi\Psi a} - \boldsymbol{y}) / 2\sigma^2 \right)}\ .$$

Single forward-backward iteration

- Analytic approximation:

$$\boxed{\operatorname{prox}_{\hat{g}}^{\delta/2}(\boldsymbol{a}) \approx \operatorname{soft}_{\mu\delta/2} \left(\boldsymbol{a} - \delta \boldsymbol{\Psi}^\dagger \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi\Psi a} - \boldsymbol{y}) / 2\sigma^2 \right)}\ .$$

Proximal MALA

Computing proximity operators for the synthesis case

- Recall posterior: $\pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$.
- Let $\hat{g}(\boldsymbol{x}(\boldsymbol{a})) = \hat{f}_1(\boldsymbol{a}) + \hat{f}_2(\boldsymbol{a})$, where $\boxed{\hat{f}_1(\boldsymbol{a}) = \mu\|\boldsymbol{a}\|_1}$ and $\boxed{\hat{f}_2(\boldsymbol{a}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2/2\sigma^2}$.

  Prior Likelihood
- Must solve an optimisation problem for each iteration!

$$\boxed{\mathrm{prox}_{\hat{g}}^{\delta/2}(\boldsymbol{a}) = \operatorname*{argmin}_{\boldsymbol{u}\in\mathbb{R}^L} \left\{ \mu\|\boldsymbol{u}\|_1 + \frac{\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{u}\|_2^2}{2\sigma^2} + \frac{\|\boldsymbol{u} - \boldsymbol{a}\|_2^2}{\delta} \right\}}\ .$$

- Taylor expansion at point $\boldsymbol{a}$: $\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{u}\|_2^2 \approx \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2 + 2(\boldsymbol{u}-\boldsymbol{a})^\top \boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a} - \boldsymbol{y})$.
- Then proximity operator approximated by

$$\boxed{\mathrm{prox}_{\hat{g}}^{\delta/2}(\boldsymbol{a}) \approx \mathrm{prox}_{\hat{f}_1}^{\delta/2}\left(\boldsymbol{a} - \delta\boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a} - \boldsymbol{y})/2\sigma^2\right)}\ .$$

Single forward-backward iteration

- Analytic approximation:

$$\boxed{\mathrm{prox}_{\hat{g}}^{\delta/2}(\boldsymbol{a}) \approx \mathrm{soft}_{\mu\delta/2}\left(\boldsymbol{a} - \delta\boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a} - \boldsymbol{y})/2\sigma^2\right)}\ .$$

MYULA

Moreau-Yosida approximation

- Moreau-Yosida approximation (Moreau envelope) of f:

$$\boxed{f_\lambda^{\mathsf{MY}}(\boldsymbol{x}) = \inf_{\boldsymbol{u}\in\mathbb{R}^N} f(\boldsymbol{u}) + \frac{\|\boldsymbol{u}-\boldsymbol{x}\|^2}{2\lambda}}$$

- Important properties of $f_\lambda^{\mathsf{MY}}(\boldsymbol{x})$:
  1. As $\lambda \to 0$, $f_\lambda^{\mathsf{MY}}(\boldsymbol{x}) \to f(\boldsymbol{x})$
  2. $\nabla f_\lambda^{\mathsf{MY}}(\boldsymbol{x}) = (\boldsymbol{x} - \mathrm{prox}_f^\lambda(\boldsymbol{x}))/\lambda$

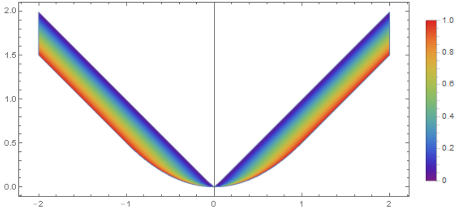


Figure: Illustration of Moreau-Yosida envelope of $|x|$ for varying λ [Credit: Stack exchange (ubpdqn)]

MYULA

Moreau-Yosida approximation

- Moreau-Yosida approximation (Moreau envelope) of f:

$$\boxed{f_\lambda^{\mathsf{MY}}(\boldsymbol{x}) = \inf_{\boldsymbol{u}\in\mathbb{R}^N} f(\boldsymbol{u}) + \frac{\|\boldsymbol{u}-\boldsymbol{x}\|^2}{2\lambda}}$$

- Important properties of $f_\lambda^{\mathsf{MY}}(\boldsymbol{x})$:

1. As $\lambda \rightarrow 0$, $f_\lambda^{\mathsf{MY}}(\boldsymbol{x}) \rightarrow f(\boldsymbol{x})$
2. $\nabla f_\lambda^{\mathsf{MY}}(\boldsymbol{x}) = (\boldsymbol{x} - \mathrm{prox}_f^\lambda(\boldsymbol{x}))/\lambda$

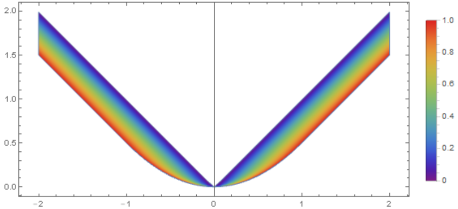


Figure: Illustration of Moreau-Yosida envelope of $|x|$ for varying λ [Credit: Stack exchange (ubpdqn)]

MYULA
MCMC sampling

Moreau-Yosida unadjusted Langevin algorithm (MYULA)
Durmus, Moulines & Pereyra (2016)

- Consider log-convex posteriors: $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$, where

$$g(\boldsymbol{x}) = \underbrace{\boxed{f_1(\boldsymbol{x})}}_{\text{Convex}} + \underbrace{\boxed{f_2(\boldsymbol{x})}}_{\text{Smooth}} .$$

- Langevin diffusion process $\mathcal{L}(t)$, with π as stationary distribution ($\mathcal{W}$ Brownian motion):

$$\mathrm{d}\mathcal{L}(t) = \frac{1}{2} \nabla \log \pi\big(\mathcal{L}(t)\big)\mathrm{d}t + \mathrm{d}\mathcal{W}(t), \quad \mathcal{L}(0) = l_0 .$$

- Euler discretisation and apply Moreau-Yosida approximation to f_1:

$$\boldsymbol{l}^{(m+1)} = \boldsymbol{l}^{(m)} + \frac{\delta}{2} \boxed{\nabla \log \pi(\boldsymbol{l}^{(m)})} + \sqrt{\delta}\boldsymbol{w}^{(m)} .$$

$$\nabla \log \pi(\boldsymbol{x}) \approx \big(\mathrm{prox}_{f_1}^{\lambda}(\boldsymbol{x}) - \boldsymbol{x}\big)/\lambda - \nabla f_2(\boldsymbol{x})$$

- No Metropolis-Hastings accept-reject step. Converges geometrically fast, where bias can be made arbitrarily small. To achieve precision target ϵ requires:
  - Worst case: order $N^5 \log^2(\epsilon^{-1})\epsilon^{-2}$ iterations.
  - Strong convexity worst case: order $N \log(N) \log^2(\epsilon^{-1})\epsilon^{-2}$ iterations.

MYULA
MCMC sampling

Moreau-Yosida unadjusted Langevin algorithm (MYULA)
Durmus, Moulines & Pereyra (2016)

- Consider log-convex posteriors: $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$, where

$$g(\boldsymbol{x}) = \underbrace{\boxed{f_1(\boldsymbol{x})}}_{\text{Convex}} + \underbrace{\boxed{f_2(\boldsymbol{x})}}_{\text{Smooth}} .$$

- Langevin diffusion process $\mathcal{L}(t)$, with π as stationary distribution ($\mathcal{W}$ Brownian motion):

$$\mathrm{d}\mathcal{L}(t) = \frac{1}{2}\nabla \log \pi\big(\mathcal{L}(t)\big)\mathrm{d}t + \mathrm{d}\mathcal{W}(t), \quad \mathcal{L}(0) = l_0 .$$

- Euler discretisation and apply Moreau-Yosida approximation to f_1:

$$\boldsymbol{l}^{(m+1)} = \boldsymbol{l}^{(m)} + \frac{\delta}{2}\boxed{\nabla \log \pi(\boldsymbol{l}^{(m)})} + \sqrt{\delta}\boldsymbol{w}^{(m)} .$$

$$\nabla \log \pi(\boldsymbol{x}) \approx \big(\mathrm{prox}_{f_1}^{\lambda}(\boldsymbol{x}) - \boldsymbol{x}\big)/\lambda - \nabla f_2(\boldsymbol{x})$$

- No Metropolis-Hastings accept-reject step. Converges geometrically fast, where bias can be made arbitrarily small. To achieve precision target ϵ requires:
  - Worst case: order $N^5 \log^2(\epsilon^{-1})\epsilon^{-2}$ iterations.
  - Strong convexity worst case: order $N \log(N) \log^2(\epsilon^{-1})\epsilon^{-2}$ iterations.

MYULA

MCMC sampling

Moreau-Yosida unadjusted Langevin algorithm (MYULA)

Durmus, Moulines & Pereyra (2016)

- Consider log-convex posteriors: $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$, where

$$g(\boldsymbol{x}) = \underbrace{\boxed{f_1(\boldsymbol{x})}}_{\text{Convex}} + \underbrace{\boxed{f_2(\boldsymbol{x})}}_{\text{Smooth}} .$$

- Langevin diffusion process $\mathcal{L}(t)$, with π as stationary distribution ($\mathcal{W}$ Brownian motion):

$$\mathrm{d}\mathcal{L}(t) = \frac{1}{2}\nabla \log \pi\big(\mathcal{L}(t)\big)\mathrm{d}t + \mathrm{d}\mathcal{W}(t), \quad \mathcal{L}(0) = l_0 .$$

- Euler discretisation and apply Moreau-Yosida approximation to f_1:

$$\boldsymbol{l}^{(m+1)} = \boldsymbol{l}^{(m)} + \frac{\delta}{2}\boxed{\nabla \log \pi(\boldsymbol{l}^{(m)})} + \sqrt{\delta}\boldsymbol{w}^{(m)} .$$

$$\nabla \log \pi(\boldsymbol{x}) \approx \big(\mathrm{prox}_{f_1}^{\lambda}(\boldsymbol{x}) - \boldsymbol{x}\big)/\lambda - \nabla f_2(\boldsymbol{x})$$

- No Metropolis-Hastings accept-reject step. Converges geometrically fast, where bias can be made arbitrarily small. To achieve precision target ϵ requires:
    - Worst case: order $N^5 \log^2(\epsilon^{-1})\epsilon^{-2}$ iterations.
    - Strong convexity worst case: order $N \log(N) \log^2(\epsilon^{-1})\epsilon^{-2}$ iterations.

MYULA

MCMC sampling

Moreau-Yosida unadjusted Langevin algorithm (MYULA)

Durmus, Moulines & Pereyra (2016)

- Consider log-convex posteriors: $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$, where

$$g(\boldsymbol{x}) = \underbrace{\boxed{f_1(\boldsymbol{x})}}_{\text{Convex}} + \underbrace{\boxed{f_2(\boldsymbol{x})}}_{\text{Smooth}} \,.$$

- Langevin diffusion process $\mathcal{L}(t)$, with π as stationary distribution ($\mathcal{W}$ Brownian motion):

$$\mathrm{d}\mathcal{L}(t) = \frac{1}{2}\nabla \log \pi\big(\mathcal{L}(t)\big)\mathrm{d}t + \mathrm{d}\mathcal{W}(t), \quad \mathcal{L}(0) = l_0 \,.$$

- Euler discretisation and apply Moreau-Yosida approximation to f_1:

$$\boldsymbol{l}^{(m+1)} = \boldsymbol{l}^{(m)} + \frac{\delta}{2}\boxed{\nabla \log \pi(\boldsymbol{l}^{(m)})} + \sqrt{\delta}\boldsymbol{w}^{(m)} \,.$$

$$\nabla \log \pi(\boldsymbol{x}) \approx \big(\mathrm{prox}_{f_1}^{\lambda}(\boldsymbol{x}) - \boldsymbol{x}\big)/\lambda - \nabla f_2(\boldsymbol{x})$$

- No Metropolis-Hastings accept-reject step. Converges geometrically fast, where bias can be made arbitrarily small. To achieve precision target ϵ requires:
  - Worst case: order $N^5 \log^2(\epsilon^{-1})\epsilon^{-2}$ iterations.
  - Strong convexity worst case: order $N \log(N) \log^2(\epsilon^{-1})\epsilon^{-2}$ iterations.

MYULA

Computing proximity operators for the analysis case

- Recall posterior: $\pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$.
- Let $\bar{g}(\boldsymbol{x}) = \bar{f}_1(\boldsymbol{x}) + \bar{f}_2(\boldsymbol{x})$, where $\boxed{\bar{f}_1(\boldsymbol{x}) = \mu \|\boldsymbol{\Psi}^\dagger \boldsymbol{x}\|_1}$ and $\boxed{\bar{f}_2(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 / 2\sigma^2}$.

  Prior Likelihood
- Only need to compute proximity operator of f_1, which can be computed analytically without any approximation:

$$\boxed{\mathrm{prox}_{\bar{f}_1}^{\delta/2}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{\Psi}\left(\mathrm{soft}_{\mu\delta/2}(\boldsymbol{\Psi}^\dagger \boldsymbol{x}) - \boldsymbol{\Psi}^\dagger \boldsymbol{x}\right)}\,.$$

MYULA

Computing proximity operators for the analysis case

- Recall posterior: $\pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$.
- Let $\bar{g}(\boldsymbol{x}) = \bar{f}_1(\boldsymbol{x}) + \bar{f}_2(\boldsymbol{x})$, where $\underbrace{\boxed{\bar{f}_1(\boldsymbol{x}) = \mu\|\boldsymbol{\Psi}^\dagger \boldsymbol{x}\|_1}}_{\text{Prior}}$ and $\underbrace{\boxed{\bar{f}_2(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2/2\sigma^2}}_{\text{Likelihood}}$.
- Only need to compute proximity operator of f_1, which can be computed analytically without any approximation:

$$\boxed{\mathrm{prox}_{\bar{f}_1}^{\delta/2}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{\Psi}\left(\mathrm{soft}_{\mu\delta/2}(\boldsymbol{\Psi}^\dagger \boldsymbol{x}) - \boldsymbol{\Psi}^\dagger \boldsymbol{x}\right)}\,.$$

MYULA

Computing proximity operators for the synthesis case

- Recall posterior: $\pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$.
- Let $\hat{g}(\boldsymbol{x}(\boldsymbol{a})) = \hat{f}_1(\boldsymbol{a}) + \hat{f}_2(\boldsymbol{a})$, where $\boxed{\hat{f}_1(\boldsymbol{a}) = \mu \|\boldsymbol{a}\|_1}$ and $\boxed{\hat{f}_2(\boldsymbol{a}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2 / 2\sigma^2}$.

  Prior Likelihood
- Only need to compute proximity operator of f_1, which can be computed analytically without any approximation:

$$\boxed{\mathrm{prox}_{\hat{f}_1}^{\delta/2}(\boldsymbol{a}) = \mathrm{soft}_{\mu\delta/2}(\boldsymbol{a})} .$$

MYULA

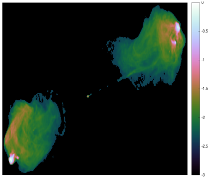
Computing proximity operators for the synthesis case

- Recall posterior: $\pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$.
- Let $\hat{g}(\boldsymbol{x}(\boldsymbol{a})) = \hat{f}_1(\boldsymbol{a}) + \hat{f}_2(\boldsymbol{a})$, where $\boxed{\hat{f}_1(\boldsymbol{a}) = \mu \|\boldsymbol{a}\|_1}$ and $\boxed{\hat{f}_2(\boldsymbol{a}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2/2\sigma^2}$.

  Prior Likelihood
- Only need to compute proximity operator of f_1, which can be computed analytically without any approximation:

$$\boxed{\mathrm{prox}_{\hat{f}_1}^{\delta/2}(\boldsymbol{a}) = \mathrm{soft}_{\mu\delta/2}(\boldsymbol{a})}\ .$$

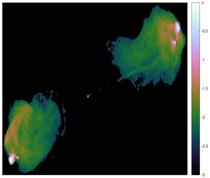
Numerical experiments
MYULA with analysis model



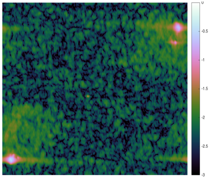
(a) Ground truth

Figure: Cygnus A

Numerical experiments
MYULA with analysis model



(a) Ground truth

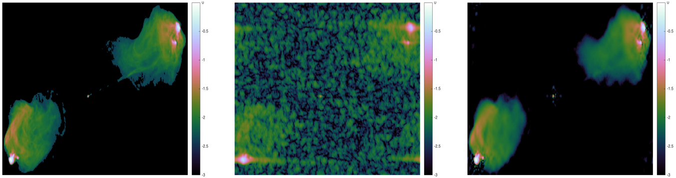


(b) Dirty image

Figure: Cygnus A

Numerical experiments

MYULA with analysis model

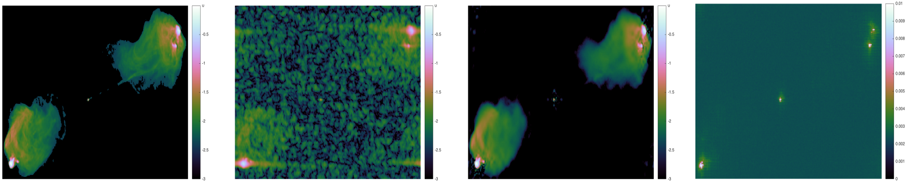


(a) Ground truth (b) Dirty image (c) Mean recovered image

Figure: Cygnus A

Numerical experiments

MYULA with analysis model

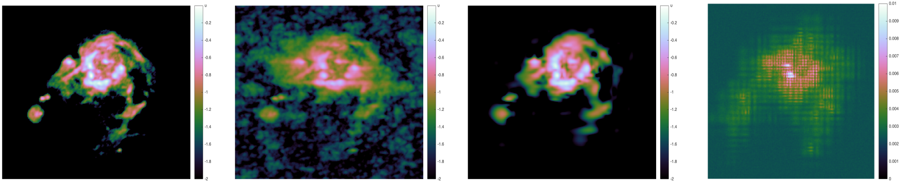


(a) Ground truth (b) Dirty image (c) Mean recovered image (d) Credible interval length

Figure: Cygnus A

Numerical experiments

MYULA with analysis model

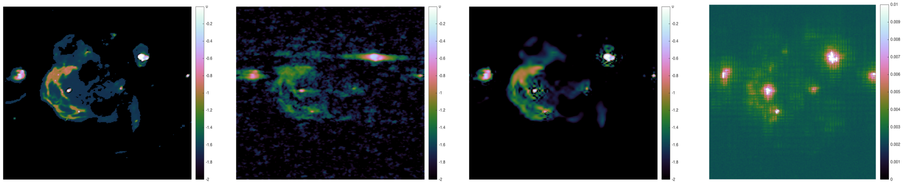


(a) Ground truth (b) Dirty image (c) Mean recovered image (d) Credible interval length

Figure: HII region of M31

Numerical experiments

MYULA with analysis model

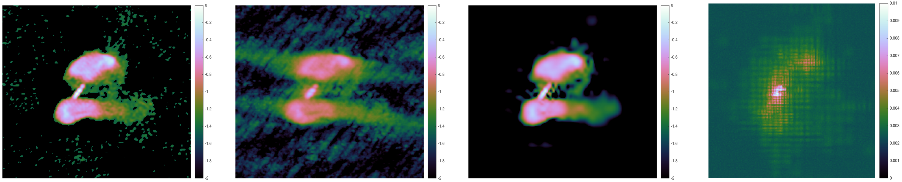


(a) Ground truth (b) Dirty image (c) Mean recovered image (d) Credible interval length

Figure: W28 Supernova remnant

Numerical experiments

MYULA with analysis model



(a) Ground truth (b) Dirty image (c) Mean recovered image (d) Credible interval length

Figure: 3C288

Numerical experiments

Computation time

Table: CPU time in minutes for Proximal MCMC sampling

Image	Method	CPU time (min) Analysis	CPU time (min) Synthesis
Cygnus A	Px-MALA	2274	1762
	MYULA	1056	942
M31	Px-MALA	1307	944
	MYULA	618	581
W28	Px-MALA	1122	879
	MYULA	646	598
3C288	Px-MALA	1144	881
	MYULA	607	538

Hypothesis testing

Method

- Perform hypothesis tests of image structure using Bayesian credible regions (Pereyra 2016b).

- Let C_α denote the highest posterior density (HPD) Bayesian credible region with confidence level $(1-\alpha)\%$ defined by posterior iso-contour: $C_\alpha = \{x : g(x) \leq \gamma_\alpha\}$.

Hypothesis testing of physical structure

1. Remove structure of interest from recovered image $x^\star$.
2. Inpaint background (noise) into region, yielding surrogate image x'.
3. Test whether $x' \in C_\alpha$:
   - If $x' \notin C_\alpha$ then reject hypothesis that structure is an artifact with confidence $(1-\alpha)\%$, i.e.
   - If $x' \in C_\alpha$ uncertainty too high to draw strong conclusions about the physical nature of the structure.

Hypothesis testing
Method

- Perform hypothesis tests of image structure using Bayesian credible regions (Pereyra 2016b).
- Let C_α denote the highest posterior density (HPD) Bayesian credible region with confidence level $(1-\alpha)\%$ defined by posterior iso-contour: $C_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$.

Hypothesis testing of physical structure

1. Remove structure of interest from recovered image $x^\star$.
2. Inpaint background (noise) into region, yielding surrogate image x'.
3. Test whether $x' \in C_\alpha$:
   - If $x' \notin C_\alpha$ then reject hypothesis that structure is an artifact with confidence $(1-\alpha)\%$, i.e.
   - If $x' \in C_\alpha$ uncertainty too high to draw strong conclusions about the physical nature of the structure.

Hypothesis testing
Method

- Perform hypothesis tests of image structure using Bayesian credible regions (Pereyra 2016b).
- Let C_α denote the highest posterior density (HPD) Bayesian credible region with confidence level $(1-\alpha)\%$ defined by posterior iso-contour: $C_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$.

Hypothesis testing of physical structure

1. Remove structure of interest from recovered image $\boldsymbol{x}^\star$.
2. Inpaint background (noise) into region, yielding surrogate image $\boldsymbol{x}'$.
3. Test whether $\boldsymbol{x}' \in C_\alpha$:
   - If $\boldsymbol{x}' \notin C_\alpha$ then reject hypothesis that structure is an artifact with confidence $(1-\alpha)\%$, *i.e.* structure most likely physical.
   - If $\boldsymbol{x}' \in C_\alpha$ uncertainly too high to draw strong conclusions about the physical nature of the structure.

Hypothesis testing
Method

- Perform hypothesis tests of image structure using Bayesian credible regions (Pereyra 2016b).
- Let C_α denote the highest posterior density (HPD) Bayesian credible region with confidence level $(1-\alpha)\%$ defined by posterior iso-contour: $C_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$.

Hypothesis testing of physical structure

1. Remove structure of interest from recovered image $\boldsymbol{x}^\star$.
2. Inpaint background (noise) into region, yielding surrogate image $\boldsymbol{x}'$.
3. Test whether $\boldsymbol{x}' \in C_\alpha$:
   - If $\boldsymbol{x}' \notin C_\alpha$ then reject hypothesis that structure is an artifact with confidence $(1-\alpha)\%$, *i.e.* structure most likely physical.
   - If $\boldsymbol{x}' \in C_\alpha$ uncertainly too high to draw strong conclusions about the physical nature of the structure.

Hypothesis testing

Method

- Perform hypothesis tests of image structure using Bayesian credible regions (Pereyra 2016b).
- Let C_α denote the highest posterior density (HPD) Bayesian credible region with confidence level $(1-\alpha)\%$ defined by posterior iso-contour: $C_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$.

Hypothesis testing of physical structure

1. Remove structure of interest from recovered image $\boldsymbol{x}^\star$.
2. Inpaint background (noise) into region, yielding surrogate image $\boldsymbol{x}'$.
3. Test whether $\boldsymbol{x}' \in C_\alpha$:
   - If $x' \notin C_\alpha$ then reject hypothesis that structure is an artifact with confidence $(1-\alpha)\%$, *i.e.* structure most likely physical.
   - If $x' \in C_\alpha$ uncertainly too high to draw strong conclusions about the physical nature of the structure.

Hypothesis testing
Method

- Perform hypothesis tests of image structure using Bayesian credible regions (Pereyra 2016b).
- Let C_α denote the highest posterior density (HPD) Bayesian credible region with confidence level $(1-\alpha)\%$ defined by posterior iso-contour: $C_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$.

Hypothesis testing of physical structure

1. Remove structure of interest from recovered image $\boldsymbol{x}^\star$.
2. Inpaint background (noise) into region, yielding surrogate image $\boldsymbol{x}'$.
3. Test whether $\boldsymbol{x}' \in C_\alpha$:
    - If $\boldsymbol{x}' \notin C_\alpha$ then reject hypothesis that structure is an artifact with confidence $(1-\alpha)\%$, *i.e.* structure most likely physical.
    - If $\boldsymbol{x}' \in C_\alpha$ uncertainly too high to draw strong conclusions about the physical nature of the structure.

Hypothesis testing
Method

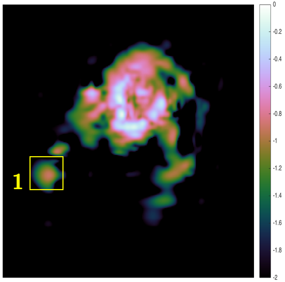
- Perform hypothesis tests of image structure using Bayesian credible regions (Pereyra 2016b).
- Let C_α denote the highest posterior density (HPD) Bayesian credible region with confidence level $(1-\alpha)\%$ defined by posterior iso-contour: $C_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$.

Hypothesis testing of physical structure

1. Remove structure of interest from recovered image $\boldsymbol{x}^\star$.
2. Inpaint background (noise) into region, yielding surrogate image $\boldsymbol{x}'$.
3. Test whether $\boldsymbol{x}' \in C_\alpha$:
    - If $\boldsymbol{x}' \notin C_\alpha$ then reject hypothesis that structure is an artifact with confidence $(1-\alpha)\%$, *i.e.* structure most likely physical.
    - If $\boldsymbol{x}' \in C_\alpha$ uncertainly too high to draw strong conclusions about the physical nature of the structure.

Hypothesis testing

Numerical experiments

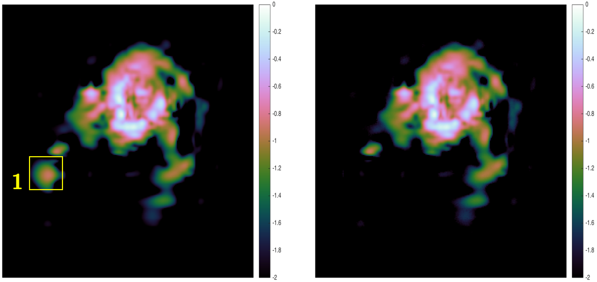


(a) Recovered image

Figure: HII region of M31

Hypothesis testing

Numerical experiments

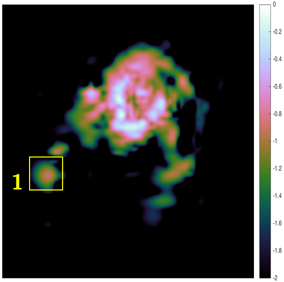


(a) Recovered image (b) Surrogate with region removed

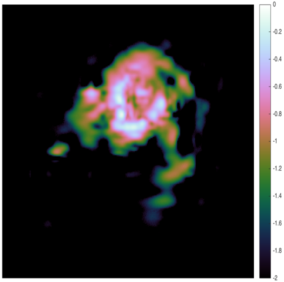
Figure: HII region of M31

Hypothesis testing

Numerical experiments



(a) Recovered image



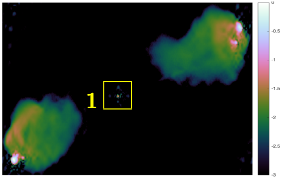
(b) Surrogate with region removed

Figure: HII region of M31

1. Reject null hypothesis

   $\Rightarrow$ structure physical

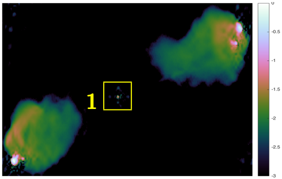
Hypothesis testing
Numerical experiments



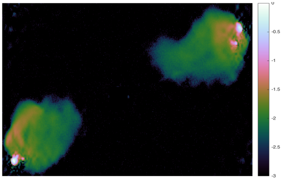
(a) Recovered image

Figure: Cygnus A

Hypothesis testing
Numerical experiments



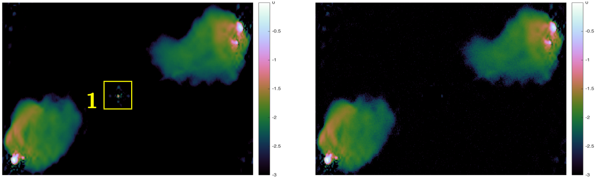
(a) Recovered image



(b) Surrogate with region removed

Figure: Cygnus A

Hypothesis testing
Numerical experiments



(a) Recovered image

(b) Surrogate with region removed

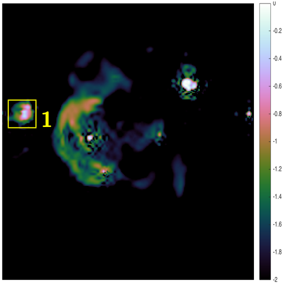
Figure: Cygnus A

1. Cannot reject null hypothesis

$\Rightarrow$ cannot make strong statistical statement about origin of structure

Hypothesis testing

Numerical experiments

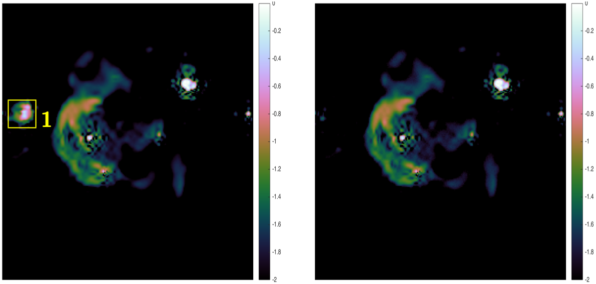


(a) Recovered image

Figure: Supernova remnant W28

Hypothesis testing

Numerical experiments

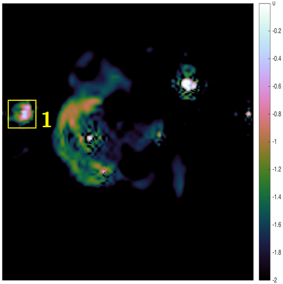


(a) Recovered image (b) Surrogate with region removed

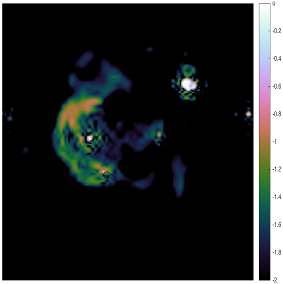
Figure: Supernova remnant W28

Hypothesis testing

Numerical experiments



(a) Recovered image



(b) Surrogate with region removed

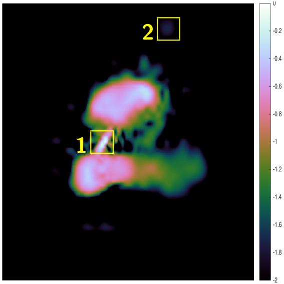
Figure: Supernova remnant W28

1. Reject null hypothesis

   ⇒ structure physical

Hypothesis testing

Numerical experiments

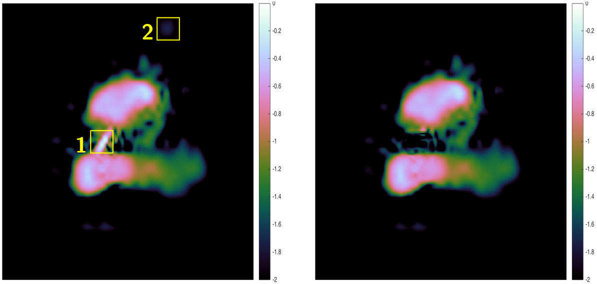


(a) Recovered image

Figure: 3C288

Hypothesis testing

Numerical experiments

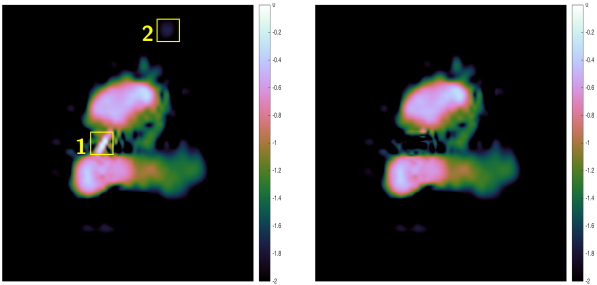


(a) Recovered image (b) Surrogate with region removed

Figure: 3C288

Hypothesis testing

Numerical experiments



(a) Recovered image

(b) Surrogate with region removed

Figure: 3C288

1. Reject null hypothesis

   ⇒ structure physical

2. Cannot reject null hypothesis

   ⇒ cannot make strong statistical statement about origin of structure

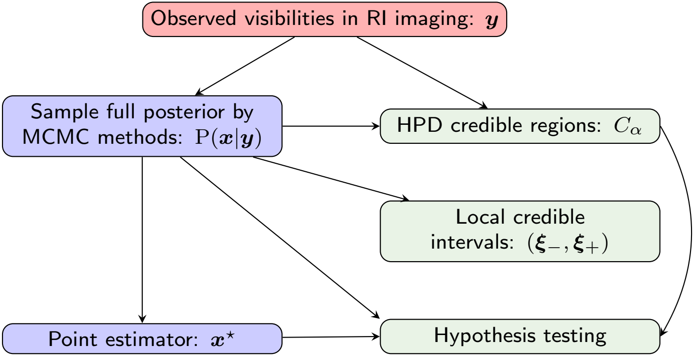
Outline

1. Radio interferometric imaging

2. Proximal MCMC sampling and uncertainty quantification

3. MAP estimation and uncertainty quantification

Proximal MCMC sampling and uncertainty quantification



Approximate Bayesian credible regions for MAP estimation

- Combine uncertainty quantification with fast sparse regularisation to scale to big-data.
- Recall C_α denotes the highest posterior density (HPD) Bayesian credible region with confidence level $(1-\alpha)\%$ defined by posterior iso-contour: $C_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$.
- Analytic approximation of γ_α:

$$\boxed{\tilde{\gamma}_\alpha = g(\boldsymbol{x}^\star) + N(\tau_\alpha + 1)}$$

where $\tau_\alpha = \sqrt{16\log(3/\alpha)/N}$ and $\alpha \in (4\mathrm{exp}(-N/3), 1)$ (Pereyra 2016b).

- Define approximate HPD regions by $\tilde{C}_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \tilde{\gamma}_\alpha\}$.
- Compute $\boldsymbol{x}^\star$ by sparse regularisation, then estimate local Bayesian credible intervals and perform hypothesis testing using approximate HPD regions.

Approximate Bayesian credible regions for MAP estimation

- Combine uncertainty quantification with fast sparse regularisation to scale to big-data.
- Recall C_α denotes the highest posterior density (HPD) Bayesian credible region with confidence level $(1-\alpha)\%$ defined by posterior iso-contour: $C_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$.
- Analytic approximation of γ_α:

$$\boxed{\tilde{\gamma}_\alpha = g(\boldsymbol{x}^\star) + N(\tau_\alpha + 1)}$$

  where $\tau_\alpha = \sqrt{16\log(3/\alpha)/N}$ and $\alpha \in (4\exp(-N/3), 1)$ (Pereyra 2016b).

- Define approximate HPD regions by $\tilde{C}_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \tilde{\gamma}_\alpha\}$.
- Compute $\boldsymbol{x}^\star$ by sparse regularisation, then estimate local Bayesian credible intervals and perform hypothesis testing using approximate HPD regions.

Approximate Bayesian credible regions for MAP estimation

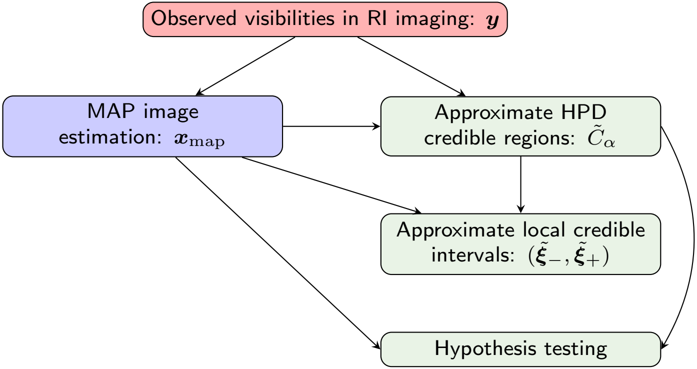
- Combine uncertainty quantification with fast sparse regularisation to scale to big-data.
- Recall C_α denotes the highest posterior density (HPD) Bayesian credible region with confidence level $(1-\alpha)\%$ defined by posterior iso-contour: $C_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$.
- Analytic approximation of γ_α:

$$\boxed{\tilde{\gamma}_\alpha = g(\boldsymbol{x}^\star) + N(\tau_\alpha + 1)}$$

where $\tau_\alpha = \sqrt{16\log(3/\alpha)/N}$ and $\alpha \in (4\mathrm{exp}(-N/3), 1)$ (Pereyra 2016b).

- Define approximate HPD regions by $\tilde{C}_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \tilde{\gamma}_\alpha\}$.
- Compute $\boldsymbol{x}^\star$ by sparse regularisation, then estimate local Bayesian credible intervals and perform hypothesis testing using approximate HPD regions.

MAP estimation and uncertainty quantification



Local Bayesian credible intervals for MAP estimation

Local Bayesian credible intervals for sparse reconstruction
(Cai, Pereyra & McEwen 2017b)

Let Ω define the area (or pixel) over which to compute the credible interval $(\tilde{\xi}_{-}, \tilde{\xi}_{+})$ and $\boldsymbol{\zeta}$ be an index vector describing Ω (*i.e.* $\zeta_i = 1$ if $i \in \Omega$ and 0 otherwise).

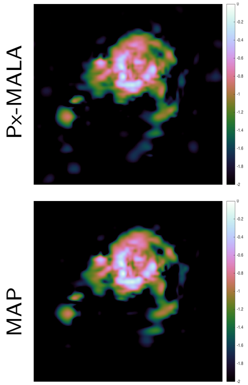
Given $\tilde{\gamma}_\alpha$ and $\boldsymbol{x}^\star$, compute the credible interval by

$$\boxed{\begin{aligned} \tilde{\xi}_{-} &= \min_{\xi} \left\{ \xi \mid g_{\boldsymbol{y}}(\boldsymbol{x}') \leq \tilde{\gamma}_\alpha, \ \forall \xi \in [-\infty, +\infty) \right\}, \\ \tilde{\xi}_{+} &= \max_{\xi} \left\{ \xi \mid g_{\boldsymbol{y}}(\boldsymbol{x}') \leq \tilde{\gamma}_\alpha, \ \forall \xi \in [-\infty, +\infty) \right\}, \end{aligned}}$$

where

$$\boxed{\boldsymbol{x}' = \boldsymbol{x}^\star (\mathcal{I} - \boldsymbol{\zeta}) + \xi \boldsymbol{\zeta}}\ .$$

Numerical experiments



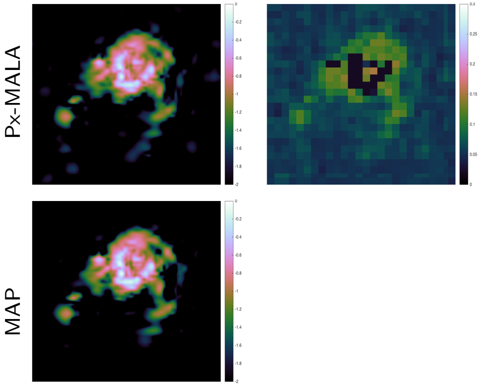
(a) point estimators

(b) local credible interval (grid size 10×10 pixels)

(c) local credible interval (grid size 20×20 pixels)

Figure: Length of local credible intervals for M31 for the analysis model.

Numerical experiments



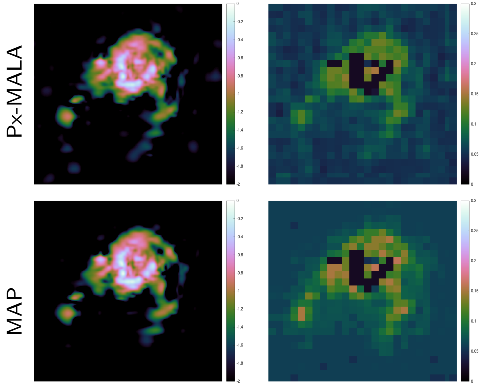
(a) point estimators

(b) local credible interval (grid size 10×10 pixels)

(c) local credible interval (grid size 20×20 pixels)

Figure: Length of local credible intervals for M31 for the analysis model.

Numerical experiments



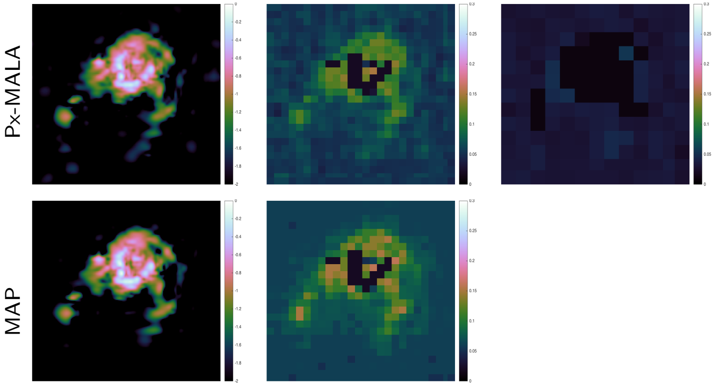
(a) point estimators

(b) local credible interval (grid size 10×10 pixels)

(c) local credible interval (grid size 20×20 pixels)

Figure: Length of local credible intervals for M31 for the analysis model.

Numerical experiments



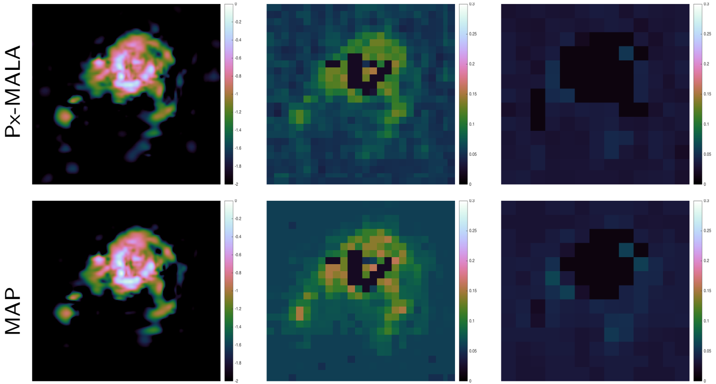
(a) point estimators

(b) local credible interval (grid size 10×10 pixels)

(c) local credible interval (grid size 20×20 pixels)

Figure: Length of local credible intervals for M31 for the analysis model.

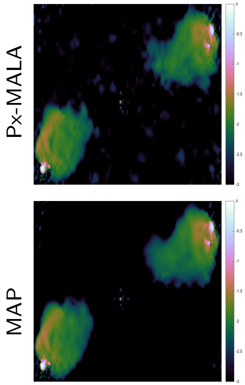
Numerical experiments



(a) point estimators (b) local credible interval (grid size 10×10 pixels) (c) local credible interval (grid size 20×20 pixels)

Figure: Length of local credible intervals for M31 for the analysis model.

Numerical experiments



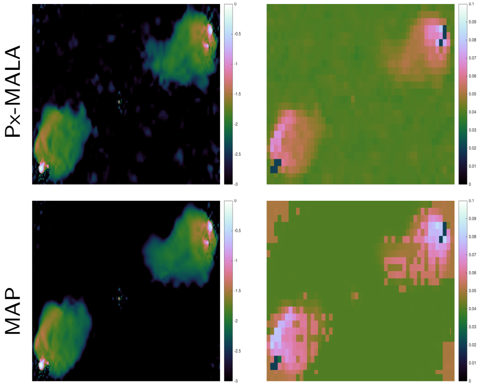
(a) point estimators

(b) local credible interval (grid size 10×10 pixels)

(c) local credible interval (grid size 20×20 pixels)

Figure: Length of local credible intervals for Cygnus A for the analysis model.

Numerical experiments



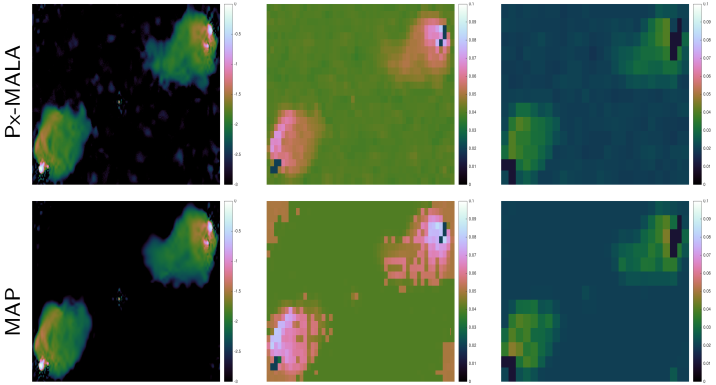
(a) point estimators

(b) local credible interval (grid size 10×10 pixels)

(c) local credible interval (grid size 20×20 pixels)

Figure: Length of local credible intervals for Cygnus A for the analysis model.

Numerical experiments



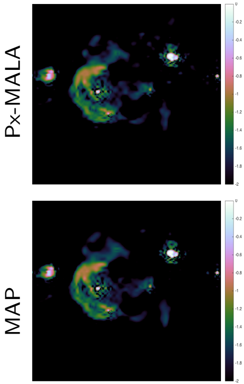
(a) point estimators

(b) local credible interval (grid size 10×10 pixels)

(c) local credible interval (grid size 20×20 pixels)

Figure: Length of local credible intervals for Cygnus A for the analysis model.

Numerical experiments



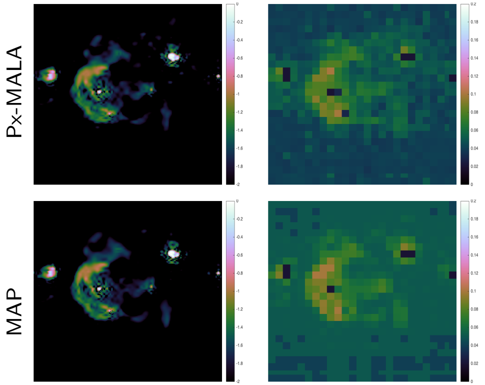
(a) point estimators

(b) local credible interval (grid size 10×10 pixels)

(c) local credible interval (grid size 20×20 pixels)

Figure: Length of local credible intervals for W28 for the analysis model.

Numerical experiments



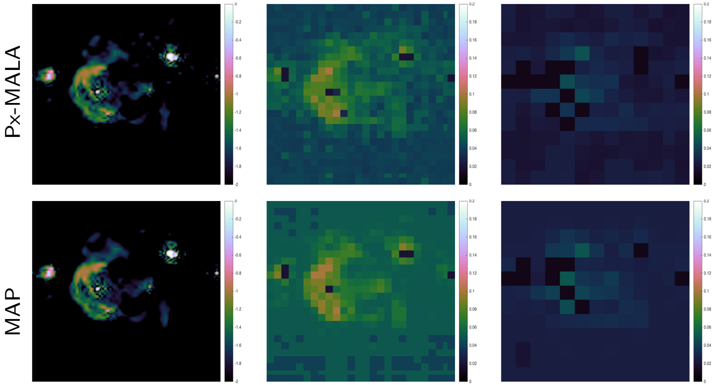
(a) point estimators

(b) local credible interval (grid size 10×10 pixels)

(c) local credible interval (grid size 20×20 pixels)

Figure: Length of local credible intervals for W28 for the analysis model.

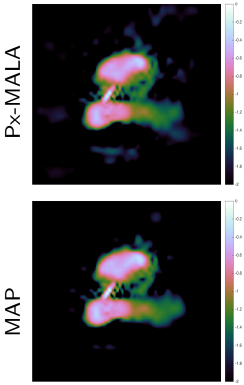
Numerical experiments



(a) point estimators (b) local credible interval (grid size 10×10 pixels) (c) local credible interval (grid size 20×20 pixels)

Figure: Length of local credible intervals for W28 for the analysis model.

Numerical experiments



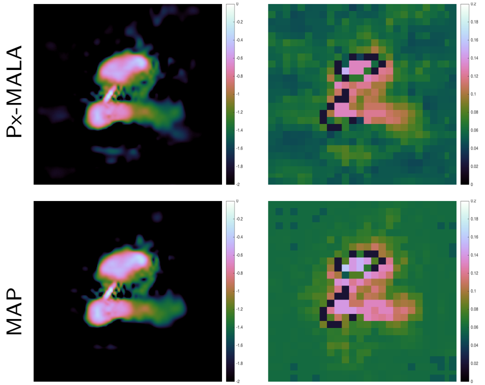
(a) point estimators

(b) local credible interval (grid size 10×10 pixels)

(c) local credible interval (grid size 20×20 pixels)

Figure: Length of local credible intervals for 3C288 for the analysis model.

Numerical experiments



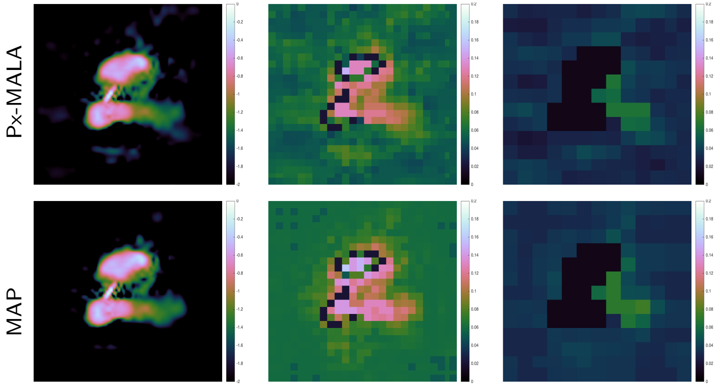
(a) point estimators

(b) local credible interval (grid size 10×10 pixels)

(c) local credible interval (grid size 20×20 pixels)

Figure: Length of local credible intervals for 3C288 for the analysis model.

Numerical experiments



(a) point estimators

(b) local credible interval (grid size 10×10 pixels)

(c) local credible interval (grid size 20×20 pixels)

Figure: Length of local credible intervals for 3C288 for the analysis model.

Computation time

Table: CPU time in minutes for Proximal MCMC sampling and MAP estimation

Image	Method	CPU time (min) Analysis	CPU time (min) Synthesis
M31	Px-MALA	1307	944
	MYULA	618	581
	MAP	.03	.02
Cygnus A	Px-MALA	2274	1762
	MYULA	1056	942
	MAP	.07	.04
W28	Px-MALA	1122	879
	MYULA	646	598
	MAP	.06	.04
3C288	Px-MALA	1144	881
	MYULA	607	538
	MAP	.03	.02

Hypothesis testing

Comparison of numerical experiments

Table: Comparison of hypothesis tests for different methods for the analysis model.

Image	Test area	Ground truth	Method	Hypothesis test
M31	1	✓	Px-MALA	✓
			MYULA	✓
			MAP	✓
Cygnus A	1	✓	Px-MALA	✗
			MYULA*	✗
			MAP	✗
W28	1	✓	Px-MALA	✓
			MYULA	✓
			MAP	✓
3C288	1	✓	Px-MALA	✓
			MYULA	✓
			MAP	✓
	2	✗	Px-MALA	✗
			MYULA	✗
			MAP	✗

(* Can correctly detect physical structure if use median point estimator.)

Conclusions

1. Sparsity-promoting priors shown to be highly effective and scalable to big-data.
   - PURIFY code provides robust framework for imaging interferometric observations (http://basp-group.github.io/purify/).
   - SOPT code for distributed sparse regularisation (http://basp-group.github.io/sopt/).

2. Proximal MCMC sampling can support sparsity-promoting priors in full Bayesian framework:
   - Recover Bayesian credible intervals.
   - Perform hypothesis testing to test whether structure physical.

3. MAP estimation (sparse regularisation) with approximate uncertainty quantification:
   - Recover Bayesian credible intervals.
   - Perform hypothesis testing to test whether structure physical.

Scalable to big-data (computational time saving $\sim 10^5$)

Supported by:

EPSRC Engineering and Physical Sciences Research Council

Science & Technology Facilities Council

Conclusions

1. Sparsity-promoting priors shown to be highly effective and scalable to big-data.
   - PURIFY code provides robust framework for imaging interferometric observations (http://basp-group.github.io/purify/).
   - SOPT code for distributed sparse regularisation (http://basp-group.github.io/sopt/).
2. Proximal MCMC sampling can support sparsity-promoting priors in full Bayesian framework:
   - Recover Bayesian credible intervals.
   - Perform hypothesis testing to test whether structure physical.
3. MAP estimation (sparse regularisation) with approximate uncertainty quantification:
   - Recover Bayesian credible intervals.
   - Perform hypothesis testing to test whether structure physical.

Scalable to big-data (computational time saving $\sim 10^5$)

Supported by:

EPSRC Engineering and Physical Sciences Research Council

Science & Technology Facilities Council

Conclusions

1. Sparsity-promoting priors shown to be highly effective and scalable to big-data.
    - PURIFY code provides robust framework for imaging interferometric observations (http://basp-group.github.io/purify/).
    - SOPT code for distributed sparse regularisation (http://basp-group.github.io/sopt/).
2. Proximal MCMC sampling can support sparsity-promoting priors in full Bayesian framework:
    - Recover Bayesian credible intervals.
    - Perform hypothesis testing to test whether structure physical.
3. MAP estimation (sparse regularisation) with approximate uncertainty quantification:
    - Recover Bayesian credible intervals.
    - Perform hypothesis testing to test whether structure physical.

Scalable to big-data (computational time saving $\sim 10^5$)

Supported by:

EPSRC
Engineering and Physical Sciences Research Council

Science & Technology Facilities Council

Conclusions

1. Sparsity-promoting priors shown to be highly effective and scalable to big-data.
    - PURIFY code provides robust framework for imaging interferometric observations (http://basp-group.github.io/purify/).
    - SOPT code for distributed sparse regularisation (http://basp-group.github.io/sopt/).

2. Proximal MCMC sampling can support sparsity-promoting priors in full Bayesian framework:
    - Recover Bayesian credible intervals.
    - Perform hypothesis testing to test whether structure physical.

3. MAP estimation (sparse regularisation) with approximate uncertainty quantification:
    - Recover Bayesian credible intervals.
    - Perform hypothesis testing to test whether structure physical.

Scalable to big-data (computational time saving $\sim 10^5$)

Supported by:

EPSRC Engineering and Physical Sciences Research Council

Science & Technology Facilities Council

Extra Slides

Analysis vs synthesis
Bayesian interpretations

Distribution and parallelisation
PURIFY reconstructions

Extra Slides

Analysis vs synthesis

Analysis vs synthesis

- Typically sparsity assumption is justified by analysing example signals in terms of atoms of the dictionary.
- Different to synthesising signals from atoms.
- Suggests an analysis-based framework (Elad *et al.* 2007, Nam *et al.* 2012):

$$\boxed{\boldsymbol{x}^\star = \underset{\boldsymbol{x}}{\arg\min}\, \|\boldsymbol{\Omega}\boldsymbol{x}\|_1 \;\text{ subject to }\; \|\boldsymbol{y} - \Phi\boldsymbol{x}\|_2 \leq \epsilon\,.}$$

analysis

- Contrast with synthesis-based approach:

$$\boxed{\boldsymbol{x}^\star = \Psi \cdot \underset{\boldsymbol{\alpha}}{\arg\min}\, \|\boldsymbol{\alpha}\|_1 \;\text{ subject to }\; \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon\,.}$$

synthesis

- For orthogonal bases $\boldsymbol{\Omega} = \Psi^\dagger$ and the two approaches are identical.

Analysis vs synthesis

Comparison

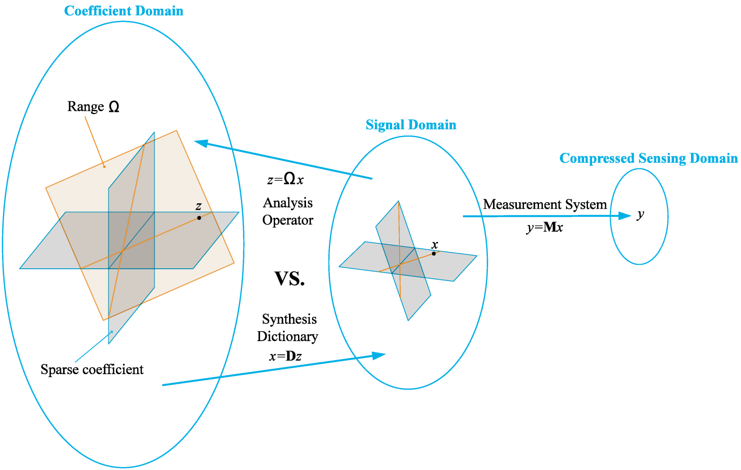


Figure: Analysis- and synthesis-based approaches **[Credit: Nam *et al.* (2012)]**.

Analysis vs synthesis
Comparison

- Synthesis-based approach is more general, while analysis-based approach more restrictive.
- More restrictive analysis-based approach may make it more robust to noise.
- The greater descriptive power of the synthesis-based approach may provide better signal representations (too descriptive?).

Extra Slides

Bayesian interpretations

Bayesian interpretations

One Bayesian interpretation of the synthesis-based approach

- Consider the inverse problem:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha} + \boldsymbol{n} \,.$$

- Assume Gaussian noise, yielding the likelihood:

$$\mathrm{P}(\boldsymbol{y} \,|\, \boldsymbol{\alpha}) \propto \exp\Big(\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha}\|_2^2 / (2\sigma^2) \Big) \,.$$

- Consider the Laplacian prior:

$$\mathrm{P}(\boldsymbol{\alpha}) \propto \exp\Big(-\beta \|\boldsymbol{\alpha}\|_1 \Big) \,.$$

- The maximum *a-posteriori* (MAP) estimate (with $\lambda = 2\beta\sigma^2$) is

$$\boxed{\boldsymbol{x}^\star_{\text{MAP-synthesis}} = \Psi \cdot \underset{\boldsymbol{\alpha}}{\arg\max}\, \mathrm{P}(\boldsymbol{\alpha} \,|\, \boldsymbol{y}) = \Psi \cdot \underset{\boldsymbol{\alpha}}{\arg\min}\, \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \,.}$$

synthesis

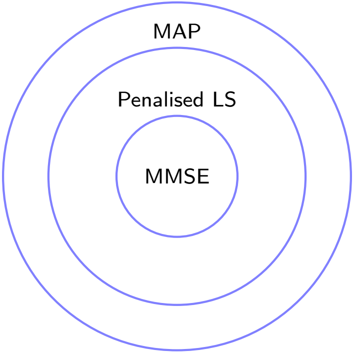
- One possible Bayesian interpretation!

- Signal may be ℓ_0-sparse, then solving ℓ_1 problem finds the correct ℓ_0-sparse solution!

Bayesian interpretations

Other Bayesian interpretations of the synthesis-based approach

- Other Bayesian interpretations are also possible (Gribonval 2011).
- Minimum mean square error (MMSE) estimators
  - $\subset$ synthesis-based estimators with appropriate penalty function, *i.e.* penalised least-squares (LS)
  - $\subset$ MAP estimators



Bayesian interpretations

One Bayesian interpretation of the analysis-based approach

- Analysis-based MAP estimate is

$$\boxed{\boldsymbol{x}^{\star}_{\text{MAP-analysis}} = \boldsymbol{\Omega}^{\dagger} \cdot \underset{\boldsymbol{\gamma} \in \textbf{column space } \boldsymbol{\Omega}}{\text{arg min}} \|\boldsymbol{y} - \Phi \boldsymbol{\Omega}^{\dagger} \boldsymbol{\gamma}\|_2^2 + \lambda \|\boldsymbol{\gamma}\|_1 \,.}$$

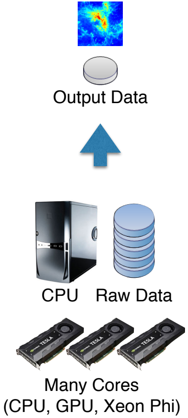
analysis

- Different to synthesis-based approach if analysis operator $\boldsymbol{\Omega}$ is not an orthogonal basis.
- Analysis-based approach more restrictive than synthesis-based.
- Similar ideas promoted by Maisinger, Hobson & Lasenby (2004) in a Bayesian framework for wavelet MEM (maximum entropy method).

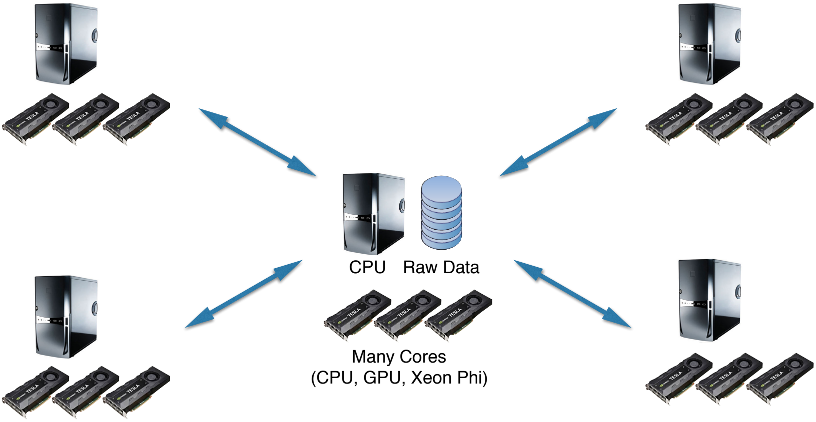
Extra Slides

Distribution and parallelisation

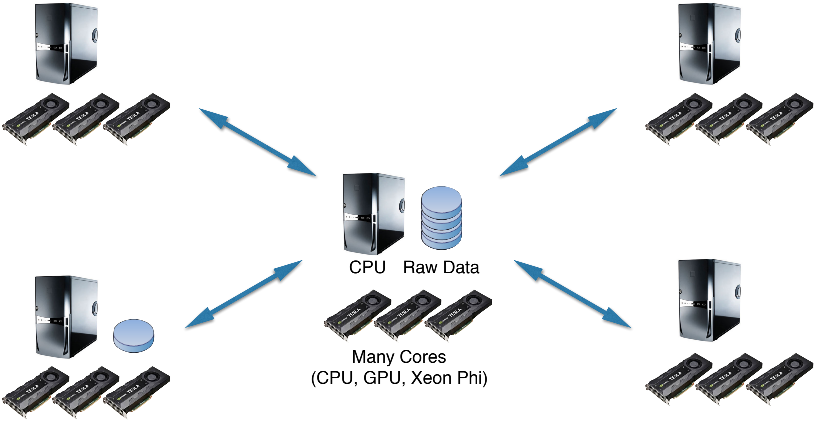
Standard algorithms



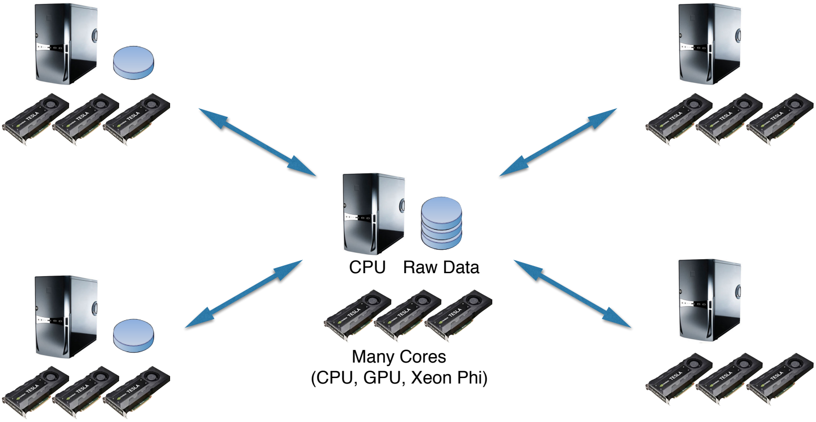
Highly distributed and parallelised algorithms



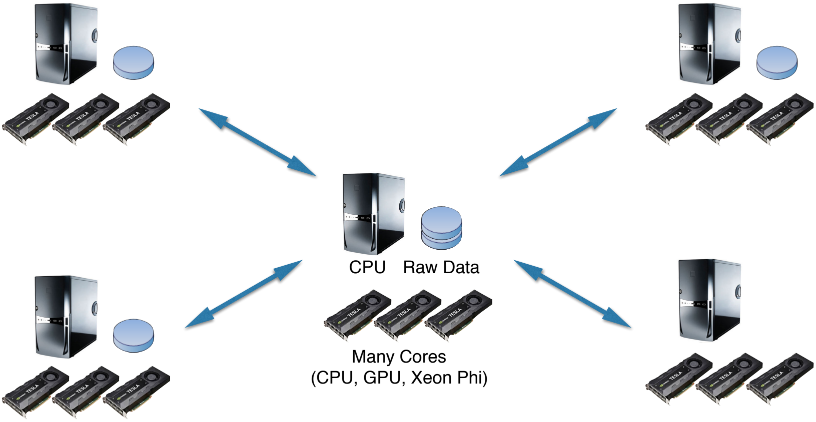
Highly distributed and parallelised algorithms



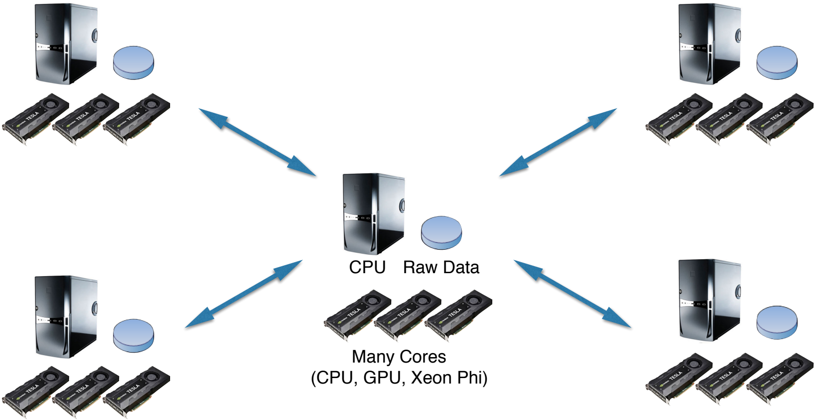
Highly distributed and parallelised algorithms



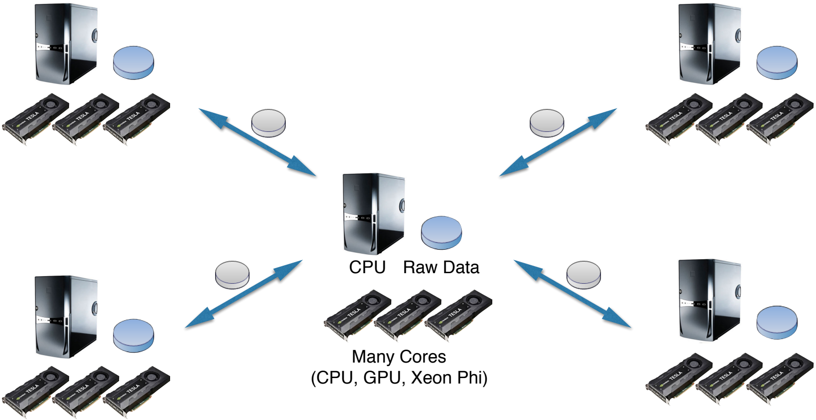
Highly distributed and parallelised algorithms



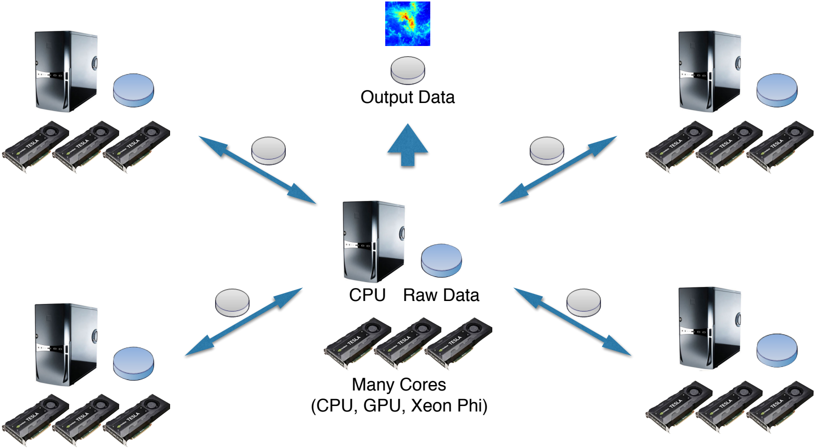
Highly distributed and parallelised algorithms



Highly distributed and parallelised algorithms



Highly distributed and parallelised algorithms



Extra Slides

PURIFY reconstructions

PURIFY reconstruction

VLA observation of 3C129

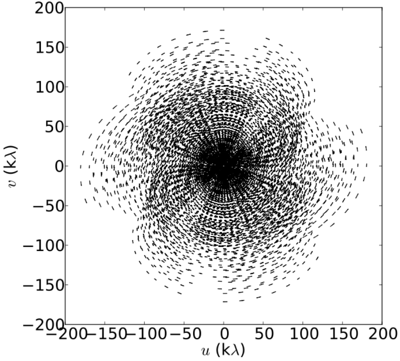
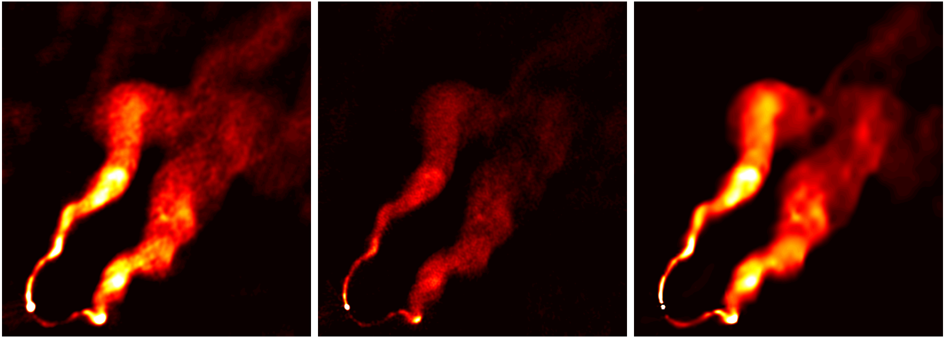


Figure: VLA visibility coverage for 3C129

PURIFY reconstruction

VLA observation of 3C129

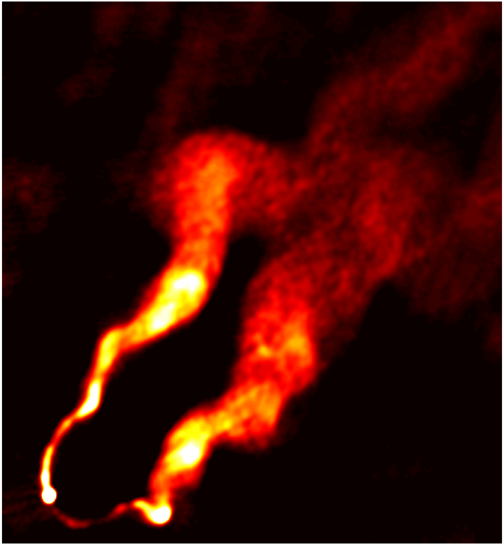


(a) CLEAN (natural) (b) CLEAN (uniform) (c) PURIFY

Figure: 3C129 recovered images (Pratley, McEwen, *et al.* 2016)

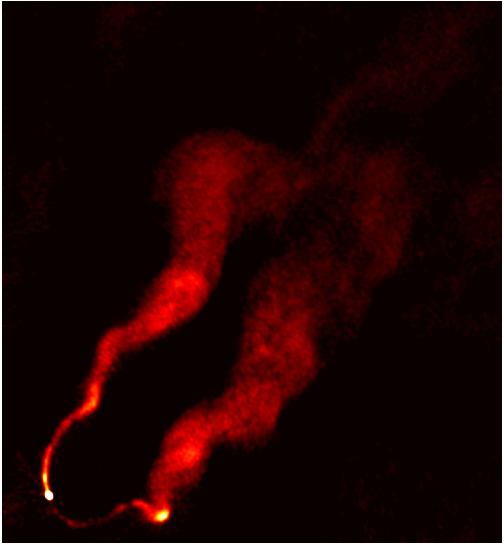
PURIFY reconstruction

VLA observation of 3C129 imaged by CLEAN (natural)



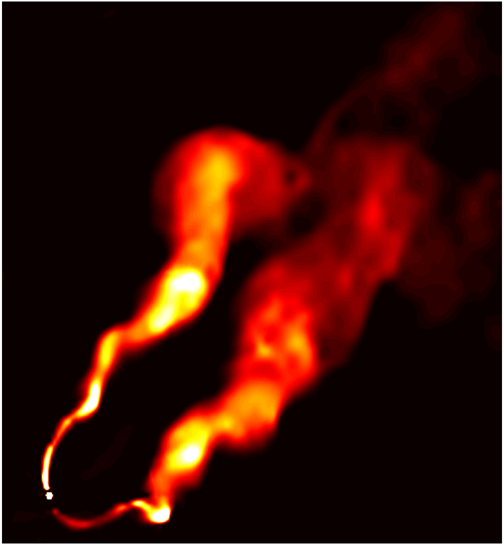
PURIFY reconstruction

VLA observation of 3C129 images by CLEAN (uniform)



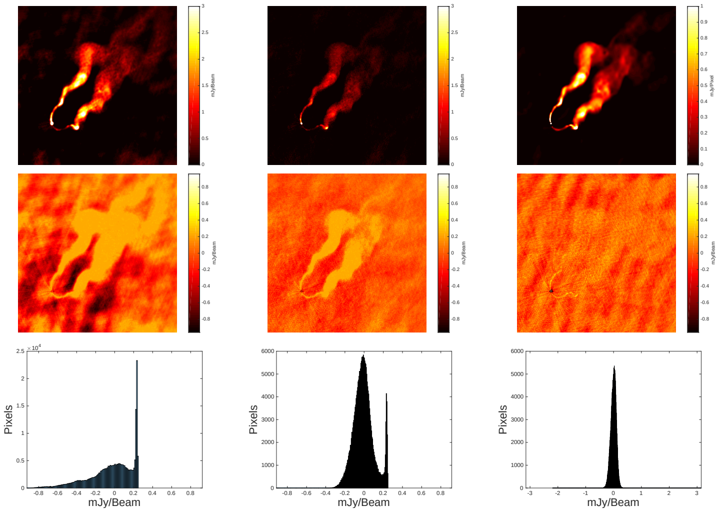
PURIFY reconstruction

VLA observation of 3C129 images by PURIFY



PURIFY reconstruction

VLA observation of 3C129



(a) CLEAN (natural) (b) CLEAN (uniform) (c) PURIFY

Figure: 3C129 recovered images and residuals (Pratley, McEwen, *et al.* 2016)

PURIFY reconstruction
VLA observation of Cygnus A

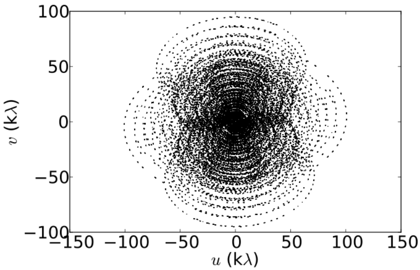
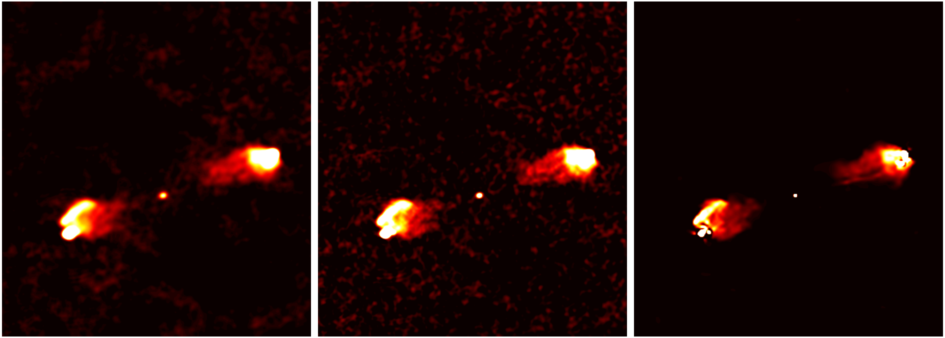


Figure: VLA visibility coverage for Cygnus A

PURIFY reconstruction

VLA observation of Cygnus A

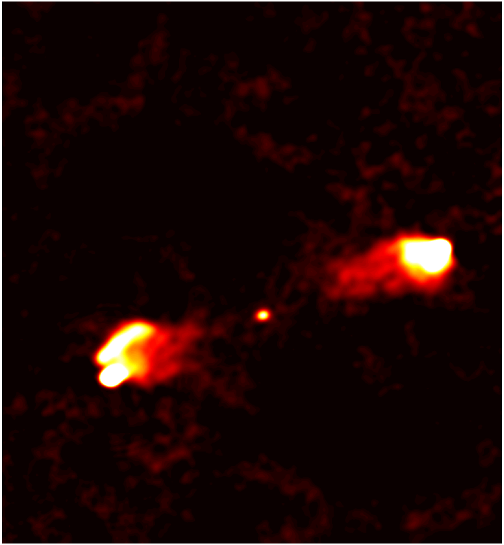


(a) CLEAN (natural) (b) CLEAN (uniform) (c) PURIFY

Figure: Cygnus A recovered images (Pratley, McEwen, *et al.* 2016)

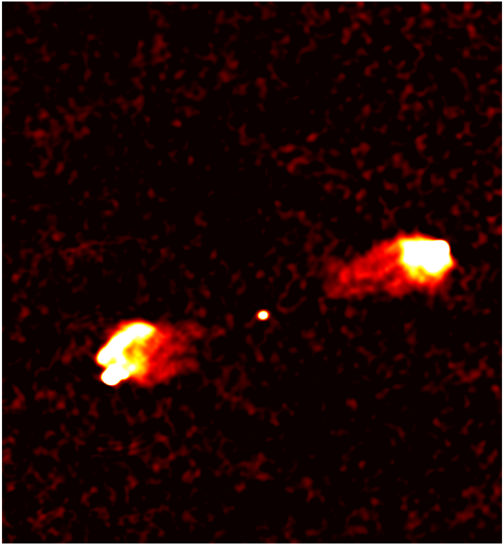
PURIFY reconstruction

VLA observation of Cygnus A imaged by CLEAN (natural)



PURIFY reconstruction

VLA observation of Cygnus A images by CLEAN (uniform)

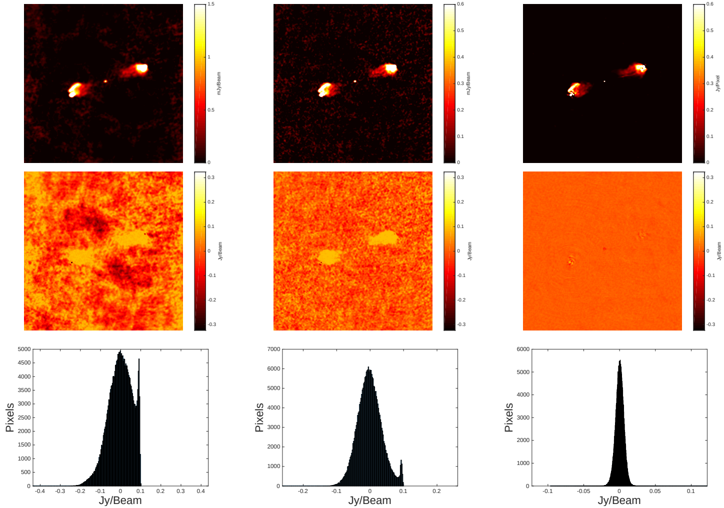


PURIFY reconstruction

VLA observation of Cygnus A images by PURIFY

PURIFY reconstruction

VLA observation of Cygnus A



(a) CLEAN (natural) (b) CLEAN (uniform) (c) PURIFY

Figure: Cygnus A recovered images and residuals (Pratley, McEwen, *et al.* 2016)

PURIFY reconstruction

ATCA observation of PKS J0334-39

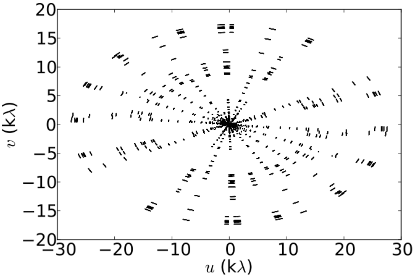
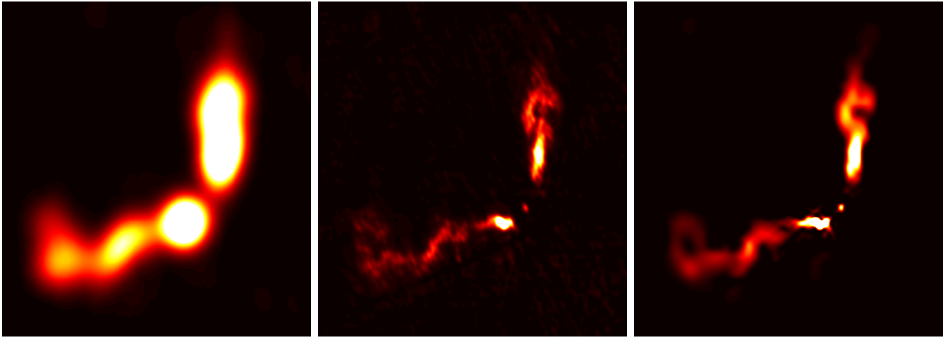


Figure: VLA visibility coverage for PKS J0334-39

PURIFY reconstruction

ATCA observation of PKS J0334-39

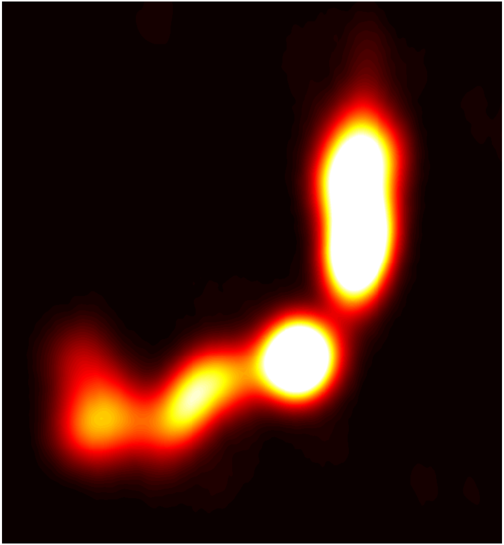


(a) CLEAN (natural) (b) CLEAN (uniform) (c) PURIFY

Figure: PKS J0334-39 recovered images (Pratley, McEwen, *et al.* 2016)

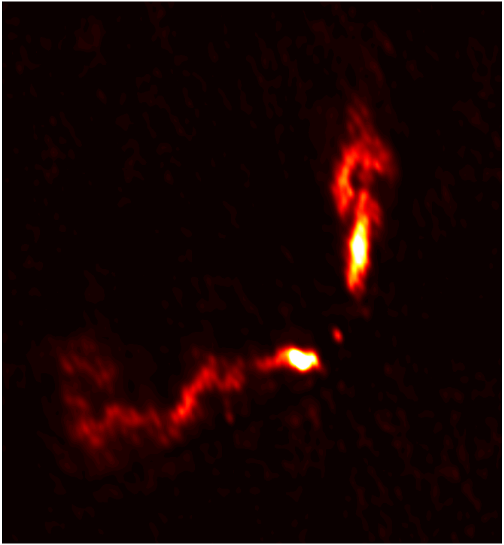
PURIFY reconstruction

VLA observation of PKS J0334-39 imaged by CLEAN (natural)



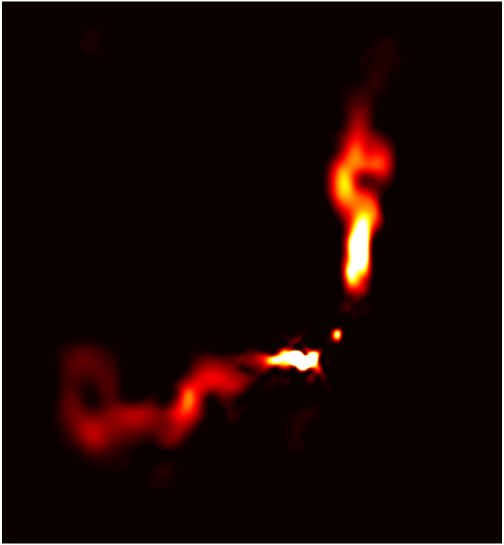
PURIFY reconstruction

VLA observation of PKS J0334-39 images by CLEAN (uniform)



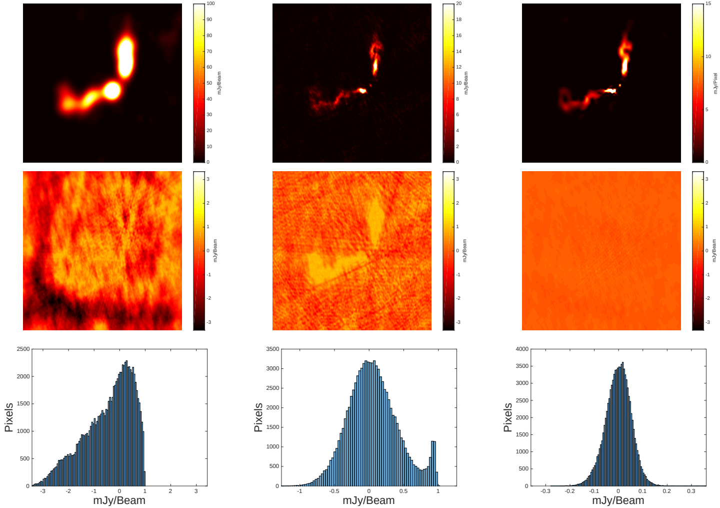
PURIFY reconstruction

VLA observation of PKS J0334-39 images by PURIFY



PURIFY reconstruction

ATCA observation of PKS J0334-39



(a) CLEAN (natural) (b) CLEAN (uniform) (c) PURIFY

Figure: PKS J0334-39 recovered images and residuals (Pratley, McEwen, *et al.* 2016)

PURIFY reconstruction

ATCA observation of PKS J0116-473

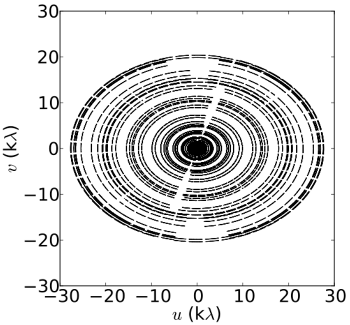
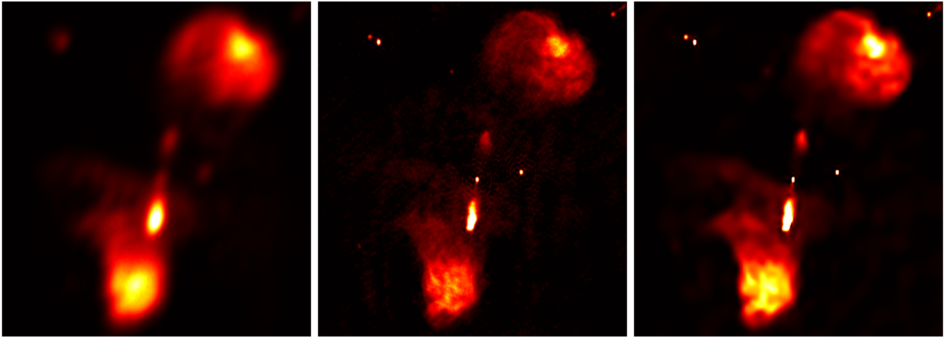


Figure: ATCA visibility coverage for Cygnus A

PURIFY reconstruction
ATCA observation of PKS J0116-473

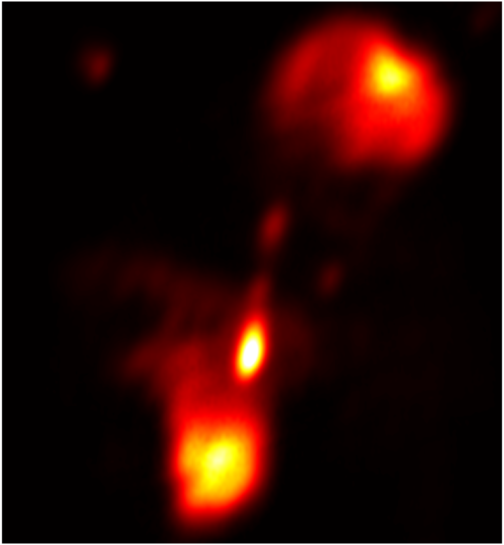


(a) CLEAN (natural) (b) CLEAN (uniform) (c) PURIFY

Figure: PKS J0116-473 recovered images (Pratley, McEwen, *et al.* 2016)

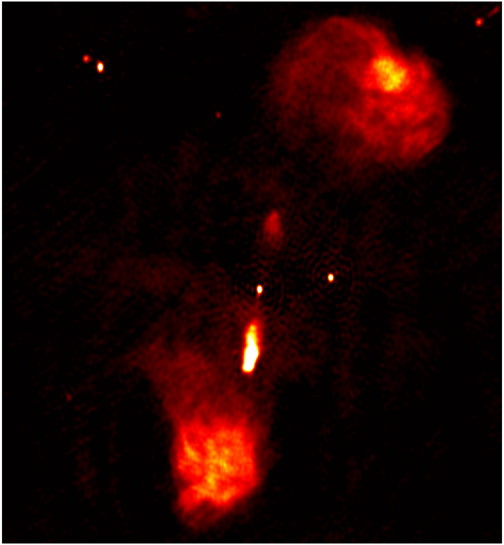
PURIFY reconstruction

VLA observation of PKS J0116-473 imaged by CLEAN (natural)



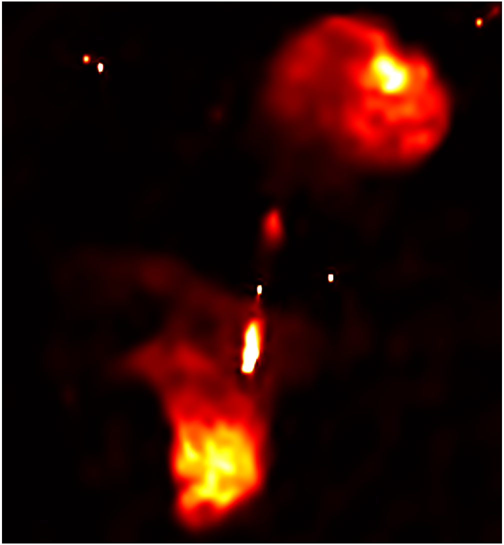
PURIFY reconstruction

VLA observation of PKS J0116-473 images by CLEAN (uniform)



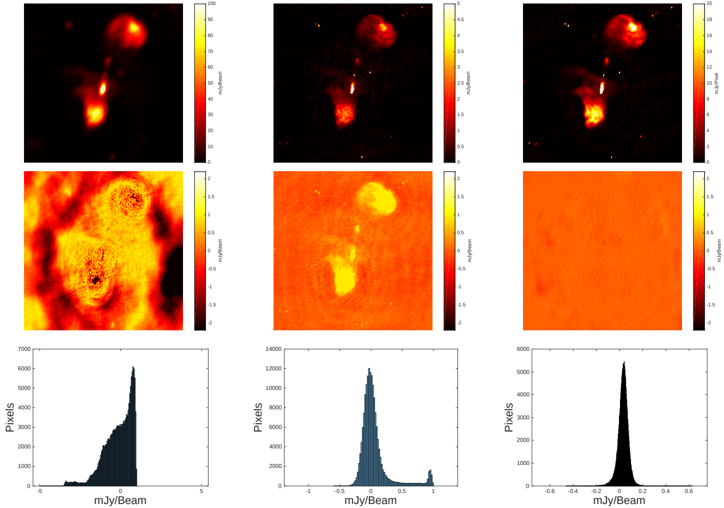
PURIFY reconstruction

VLA observation of PKS J0116-473 images by PURIFY



PURIFY reconstruction

ATCA observation of PKS J0116-473



(a) CLEAN (natural) (b) CLEAN (uniform) (c) PURIFY

Figure: PKS J0116-473 recovered images and residuals (Pratley, McEwen, *et al.* 2016)

PURIFY reconstructions

Table: Root-mean-square of residuals of each reconstruction (units in mJy/Beam)

Observation	PURIFY	CLEAN (natural)	CLEAN (uniform)
3C129	0.10	0.23	0.11
Cygnus A	6.1	59	36
PKS J0334-39	0.052	1.00	0.37
PKS J0116-473	0.054	0.88	0.24