# Statistical approaches for sparse radio interferometric imaging

## Error estimation in radio interferometry imaging
## or
## "Bayesian compressive sensing"

Jason McEwen
www.jasonmcewen.org
@jasonmcewen

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

CALIM 2016, Socorro, October 2016

# Outline

1. Bayesian interpretations of interferometric imaging techniques
2. Proximal MCMC sampling
3. Sparse regularisation with Bayesian credible regions

# Outline

# Radio interferometric inverse problem

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

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

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

# 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)...

# Radio interferometric inverse problem

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

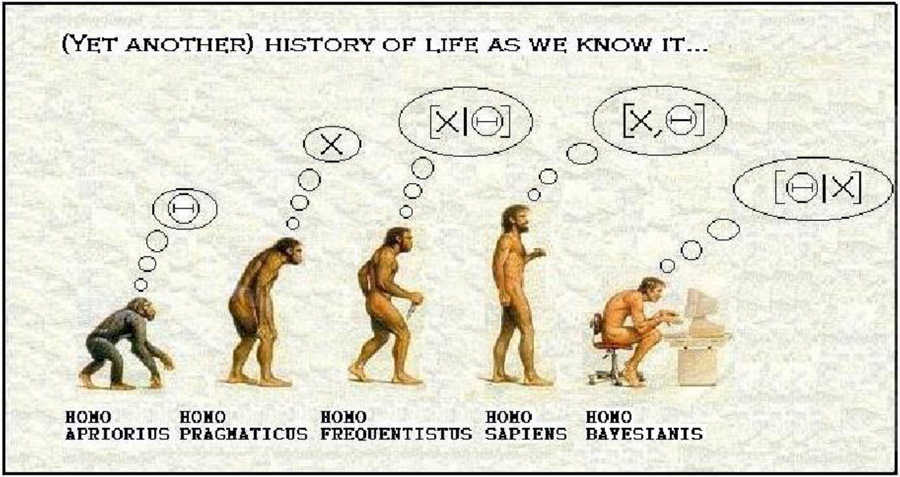
$$\boldsymbol{y} = \boldsymbol{\Phi} \boldsymbol{x} + \boldsymbol{n},$$

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.* $\boldsymbol{\Phi} = \mathbf{GFA}$, 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.

# Bayesian evolution



# Bayesian inference

- Given data $\boldsymbol{y}$ (visibilities) and model $M$ (interferometric telescope with Gaussian noise), we want a full probabilistic description of our knowledge of the underlying sky image $\boldsymbol{x}$.

- Bayes to the rescue:

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}, M) = \frac{\mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x}, M)\, \mathrm{P}(\boldsymbol{x} \mid M)}{\mathrm{P}(\boldsymbol{y} \mid M)}$$

Bayes Theorem

- Bayes theorem in words:

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

- How do we perform Bayesian inference in practice?

  ⇒ maximum a-posteriori (MAP) estimates and sampling approaches (MCMC)

  (and many others)

# Bayesian inference

- Given data $\boldsymbol{y}$ (visibilities) and model $M$ (interferometric telescope with Gaussian noise), we want a full probabilistic description of our knowledge of the underlying sky image $\boldsymbol{x}$.

- Bayes to the rescue:

$$\mathrm{P}(\boldsymbol{x} \,|\, \boldsymbol{y}, M) = \frac{\mathrm{P}(\boldsymbol{y} \,|\, \boldsymbol{x}, M)\, \mathrm{P}(\boldsymbol{x} \,|\, M)}{\mathrm{P}(\boldsymbol{y} \,|\, M)}$$

Bayes Theorem

- Bayes theorem in words:

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

- How do we perform Bayesian inference in practice?

  $\Rightarrow$ maximum a-posteriori (MAP) estimates and sampling approaches (MCMC)

  (and many others)

# Bayesian inference

- Given data $\boldsymbol{y}$ (visibilities) and model $M$ (interferometric telescope with Gaussian noise), we want a full probabilistic description of our knowledge of the underlying sky image $\boldsymbol{x}$.

- Bayes to the rescue:

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}, M) = \frac{\mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x}, M)\, \mathrm{P}(\boldsymbol{x} \mid M)}{\mathrm{P}(\boldsymbol{y} \mid M)}$$

Bayes Theorem

- Bayes theorem in words:

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

- How do we perform Bayesian inference in practice?

  $\Rightarrow$ maximum a-posteriori (MAP) estimates and sampling approaches (MCMC)

  (and many others)

# Bayes in practice

## MAP and MCMC sampling

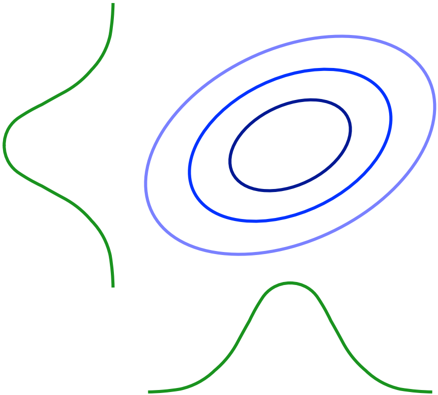


Figure: Probability distribution to explore in 2D

# Bayes in practice

## MAP and MCMC sampling

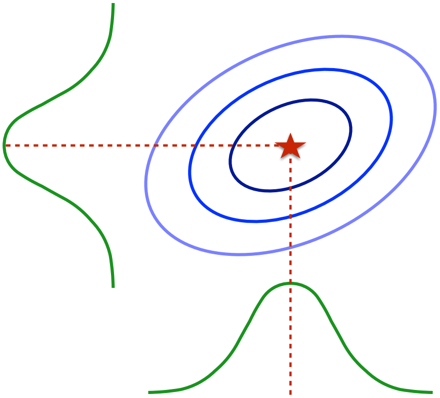


Figure: Maximum a-posteriori (MAP) estimate

# Bayes in practice

## MAP and MCMC sampling

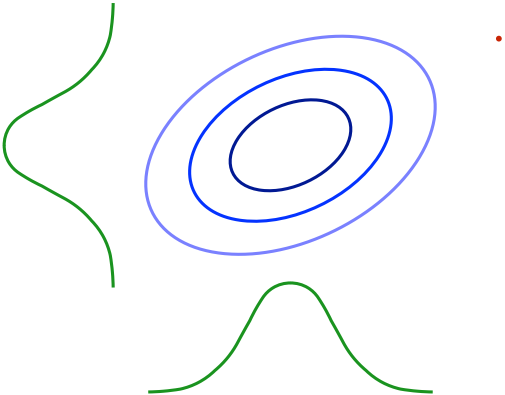


Figure: Markov Chain Monte Carlo (MCMC) sampling

# Bayes in practice
## MAP and MCMC sampling

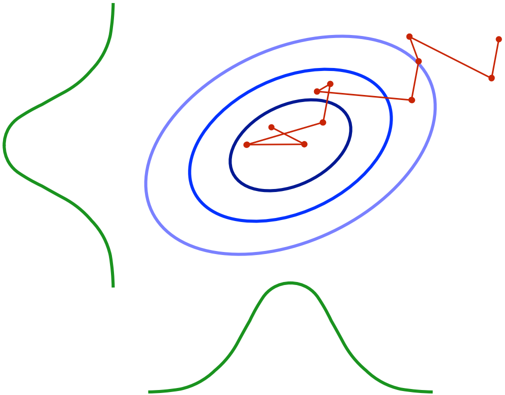


Figure: Markov Chain Monte Carlo (MCMC) sampling

# Bayes in practice

## MAP and MCMC sampling

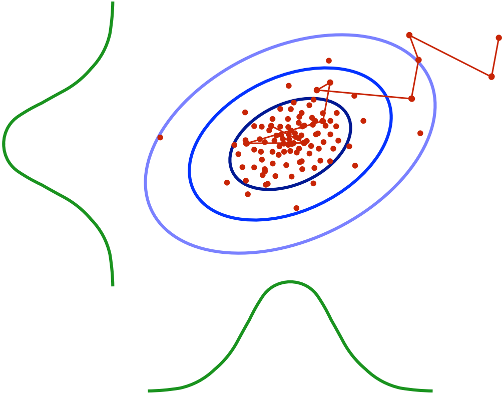


Figure: Markov Chain Monte Carlo (MCMC) sampling

# MAP estimation and regularisation

## Hint: they're the same thing!

- Many interferometric imaging approaches are based on regularisation (*i.e.* minimising an objective function comprised of a data-fidelity penalty and a regularisation penalty).
- Consider the MAP estimation problem. . .

# MAP estimation and regularisation

## Hint: they're the same thing!

- Start with Bayes Theorem (ignore normalising evidence):

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) \propto \mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x}) \mathrm{P}(\boldsymbol{x}) \,, \quad \textit{i.e.} \text{ posterior} \propto \text{likelihood} \times \text{prior}$$

- MAP estimator:

# MAP estimation and regularisation

## Hint: they're the same thing!

- Start with Bayes Theorem (ignore normalising evidence):

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) \propto \mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x}) \mathrm{P}(\boldsymbol{x}) \,, \quad \textit{i.e. } \text{posterior} \propto \text{likelihood} \times \text{prior}$$

  Define likelihood (assuming Gaussian noise) and prior:

- MAP estimator:

# MAP estimation and regularisation

## Hint: they're the same thing!

- Start with Bayes Theorem (ignore normalising evidence):

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) \propto \mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x})\mathrm{P}(\boldsymbol{x}) \,, \quad \textit{i.e.} \text{ posterior} \propto \text{likelihood} \times \text{prior}$$

Define likelihood (assuming Gaussian noise) and prior:

$$\mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x}) \propto \exp\left(-\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2/(2\sigma^2)\right)$$

Likelihood

- MAP estimator:

# MAP estimation and regularisation

## Hint: they're the same thing!

- Start with Bayes Theorem (ignore normalising evidence):

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) \propto \mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x})\mathrm{P}(\boldsymbol{x}) \,, \quad \textit{i.e.} \text{ posterior} \propto \text{likelihood} \times \text{prior}$$

Define likelihood (assuming Gaussian noise) and prior:

$$\mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x}) \propto \exp\left(-\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2/(2\sigma^2)\right)$$

Likelihood

$$\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-R(\boldsymbol{x})\right)$$

Prior

- MAP estimator:

# MAP estimation and regularisation

## Hint: they're the same thing!

- Start with Bayes Theorem (ignore normalising evidence):

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) \propto \mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x}) \mathrm{P}(\boldsymbol{x}) , \quad \textit{i.e.} \text{ posterior} \propto \text{likelihood} \times \text{prior}$$

Define likelihood (assuming Gaussian noise) and prior:

$$\mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x}) \propto \exp\left(-\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2/(2\sigma^2)\right)$$

Likelihood

$$\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-R(\boldsymbol{x})\right)$$

Prior

Consider log-posterior:

$$\log \mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = -\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2/(2\sigma^2) - R(\boldsymbol{x}) + \text{const.}$$

- MAP estimator:

# MAP estimation and regularisation

## Hint: they're the same thing!

- Start with Bayes Theorem (ignore normalising evidence):

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) \propto \mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x}) \mathrm{P}(\boldsymbol{x}) \,, \quad \textit{i.e.} \text{ posterior} \propto \text{likelihood} \times \text{prior}$$

Define likelihood (assuming Gaussian noise) and prior:

$$\mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x}) \propto \exp\Big(-\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 / (2\sigma^2)\Big)$$

Likelihood

$$\mathrm{P}(\boldsymbol{x}) \propto \exp\Big(-R(\boldsymbol{x})\Big)$$

Prior

Consider log-posterior:

$$\log \mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = -\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 / (2\sigma^2) - R(\boldsymbol{x}) + \text{const.}$$

- MAP estimator:

$$\boldsymbol{x}_{\mathrm{map}} = \arg\max_{\boldsymbol{x}} \Big[\log \mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x})\Big]$$

# MAP estimation and regularisation

## Hint: they're the same thing!

- Start with Bayes Theorem (ignore normalising evidence):

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) \propto \mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x}) \mathrm{P}(\boldsymbol{x}) \;, \quad \textit{i.e.}\text{ posterior} \propto \text{likelihood} \times \text{prior}$$

Define likelihood (assuming Gaussian noise) and prior:

$$\mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x}) \propto \exp\left(-\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2/(2\sigma^2)\right)$$

Likelihood

$$\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-R(\boldsymbol{x})\right)$$

Prior

Consider log-posterior:

$$\log \mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = -\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2/(2\sigma^2) - R(\boldsymbol{x}) + \text{const.}$$

- MAP estimator:

$$\boldsymbol{x}_{\mathrm{map}} = \arg\max_{\boldsymbol{x}} \left[\log \mathrm{P}(\boldsymbol{y} \mid \boldsymbol{x})\right] = \arg\min_{\boldsymbol{x}} \left[\underbrace{\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2}_{\text{Data fidelity}} + \underbrace{\lambda\, R(\boldsymbol{x})}_{\text{Regulariser}}\right]$$

# CLEAN and MEM as MAP estimators

- CLEAN

  Consider the sparse prior: $\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-\beta \left\|\boldsymbol{x}\right\|_0\right)$.

  Corresponding MAP estimator is:

$$\boldsymbol{x}_{\text{clean}} = \arg\min_{\boldsymbol{x}} \left[ \left\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\right\|_2^2 + \lambda \left\|\boldsymbol{x}\right\|_0 \right]$$

  (Laplace prior $\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-\beta \left\|\boldsymbol{x}\right\|_1\right)$ is good proxy; Wiaux *et al.* 2009)

- MEM

  Consider the entropic prior: $\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-\beta\, \boldsymbol{x}^\dagger \log \boldsymbol{x}\right)$.

  Corresponding MAP estimator is:

$$\boldsymbol{x}_{\text{mem}} = \arg\min_{\boldsymbol{x}} \left[ \left\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\right\|_2^2 + \lambda\, \boldsymbol{x}^\dagger \log \boldsymbol{x} \right]$$

(In practice some differences: CLEAN does not solve MAP problem exactly; MEM considered in RI imposes additional constraints.)

# CLEAN and MEM as MAP estimators

- CLEAN

  Consider the sparse prior: $\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-\beta \left\|\boldsymbol{x}\right\|_0\right)$.

  Corresponding MAP estimator is:

$$\boldsymbol{x}_{\text{clean}} = \arg\min_{\boldsymbol{x}} \left[ \left\| \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x} \right\|_2^2 + \lambda \left\|\boldsymbol{x}\right\|_0 \right]$$

(Laplace prior $\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-\beta \left\|\boldsymbol{x}\right\|_1\right)$ is good proxy; Wiaux *et al.* 2009)

- MEM

  Consider the entropic prior: $\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-\beta\, \boldsymbol{x}^\dagger \log \boldsymbol{x}\right)$.

  Corresponding MAP estimator is:

$$\boldsymbol{x}_{\text{mem}} = \arg\min_{\boldsymbol{x}} \left[ \left\| \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x} \right\|_2^2 + \lambda\, \boldsymbol{x}^\dagger \log \boldsymbol{x} \right]$$

(In practice some differences: CLEAN does not solve MAP problem exactly; MEM considered in RI imposes additional constraints.)

# CLEAN and MEM as MAP estimators

- CLEAN

  Consider the sparse prior: $\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-\beta \left\|\boldsymbol{x}\right\|_0\right)$.

  Corresponding MAP estimator is:

$$\boldsymbol{x}_{\text{clean}} = \arg\min_{\boldsymbol{x}} \left[\left\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\right\|_2^2 + \lambda \left\|\boldsymbol{x}\right\|_0\right]$$

(Laplace prior $\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-\beta \left\|\boldsymbol{x}\right\|_1\right)$ is good proxy; Wiaux *et al.* 2009)

- MEM

  Consider the entropic prior: $\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-\beta\, \boldsymbol{x}^\dagger \log \boldsymbol{x}\right)$.

  Corresponding MAP estimator is:

$$\boldsymbol{x}_{\text{mem}} = \arg\min_{\boldsymbol{x}} \left[\left\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\right\|_2^2 + \lambda\, \boldsymbol{x}^\dagger \log \boldsymbol{x}\right]$$

(In practice some differences: CLEAN does not solve MAP problem exactly;
MEM considered in RI imposes additional constraints.)

# CLEAN and MEM as MAP estimators

- CLEAN

  Consider the sparse prior: $\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-\beta \left\|\boldsymbol{x}\right\|_0\right)$.

  Corresponding MAP estimator is:

$$\boldsymbol{x}_{\text{clean}} = \arg\min_{\boldsymbol{x}} \left[ \left\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\right\|_2^2 + \lambda \left\|\boldsymbol{x}\right\|_0 \right]$$

(Laplace prior $\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-\beta \left\|\boldsymbol{x}\right\|_1\right)$ is good proxy; Wiaux *et al.* 2009)

- MEM

  Consider the entropic prior: $\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-\beta\, \boldsymbol{x}^\dagger \log \boldsymbol{x}\right)$.

  Corresponding MAP estimator is:

$$\boldsymbol{x}_{\text{mem}} = \arg\min_{\boldsymbol{x}} \left[ \left\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\right\|_2^2 + \lambda\, \boldsymbol{x}^\dagger \log \boldsymbol{x} \right]$$

(In practice some differences: CLEAN does not solve MAP problem exactly; MEM considered in RI imposes additional constraints.)

# CLEAN and MEM as MAP estimators

- CLEAN

  Consider the sparse prior: $\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-\beta \left\|\boldsymbol{x}\right\|_0\right)$.

  Corresponding MAP estimator is:

$$\boldsymbol{x}_{\text{clean}} = \arg\min_{\boldsymbol{x}} \left[ \left\| \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x} \right\|_2^2 + \lambda \left\| \boldsymbol{x} \right\|_0 \right]$$

(Laplace prior $\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-\beta \left\|\boldsymbol{x}\right\|_1\right)$ is good proxy; Wiaux *et al.* 2009)

- MEM

  Consider the entropic prior: $\mathrm{P}(\boldsymbol{x}) \propto \exp\left(-\beta\, \boldsymbol{x}^\dagger \log \boldsymbol{x}\right)$.

  Corresponding MAP estimator is:

$$\boldsymbol{x}_{\text{mem}} = \arg\min_{\boldsymbol{x}} \left[ \left\| \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x} \right\|_2^2 + \lambda\, \boldsymbol{x}^\dagger \log \boldsymbol{x} \right]$$

(In practice some differences: CLEAN does not solve MAP problem exactly;
MEM considered in RI imposes additional constraints.)

# Compressive sensing

## Synthesis framework

- Consider sparsifying representation (*e.g.* wavelet basis):

$$\boldsymbol{x} = \sum_i \Psi_i \alpha_i = \begin{pmatrix} | \\ \Psi_0 \\ | \end{pmatrix} \alpha_0 + \begin{pmatrix} | \\ \Psi_1 \\ | \end{pmatrix} \alpha_1 + \cdots \quad \Rightarrow \quad \boxed{\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}}$$

- Recover (wavelet) coefficients $\boldsymbol{\alpha}$ of image $\boldsymbol{x}$.
- Consider the Laplacian prior on coefficients: $\mathrm{P}(\boldsymbol{\alpha}) \propto \exp\left(-\beta \left\|\boldsymbol{\alpha}\right\|_1\right)$.
- Sparse synthesis regularisation problem:

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

Synthesis framework

# Compressive sensing
## Synthesis framework

- Consider sparsifying representation (*e.g.* wavelet basis):

$$\boldsymbol{x} = \sum_i \Psi_i \alpha_i = \begin{pmatrix} | \\ \Psi_0 \\ | \end{pmatrix} \alpha_0 + \begin{pmatrix} | \\ \Psi_1 \\ | \end{pmatrix} \alpha_1 + \cdots \quad \Rightarrow \quad \boxed{\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}}$$

- Recover (wavelet) coefficients $\boldsymbol{\alpha}$ of image $\boldsymbol{x}$.
- Consider the Laplacian prior on coefficients: $\mathrm{P}(\boldsymbol{\alpha}) \propto \exp\left(-\beta \left\|\boldsymbol{\alpha}\right\|_1\right)$.
- Sparse synthesis regularisation problem:

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

Synthesis framework

# Compressive sensing
## Synthesis framework

- Consider sparsifying representation (*e.g.* wavelet basis):

$$\boldsymbol{x} = \sum_i \Psi_i \alpha_i = \begin{pmatrix} | \\ \Psi_0 \\ | \end{pmatrix} \alpha_0 + \begin{pmatrix} | \\ \Psi_1 \\ | \end{pmatrix} \alpha_1 + \cdots \quad \Rightarrow \quad \boxed{\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}}$$

- Recover (wavelet) coefficients $\boldsymbol{\alpha}$ of image $\boldsymbol{x}$.
- Consider the Laplacian prior on coefficients: $\mathrm{P}(\boldsymbol{\alpha}) \propto \exp\left(-\beta \left\|\boldsymbol{\alpha}\right\|_1\right)$.
- Sparse synthesis regularisation problem:

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

Synthesis framework

# Compressive sensing

## Synthesis framework

- Consider sparsifying representation (*e.g.* wavelet basis):

$$\boldsymbol{x} = \sum_i \Psi_i \alpha_i = \begin{pmatrix} | \\ \Psi_0 \\ | \end{pmatrix} \alpha_0 + \begin{pmatrix} | \\ \Psi_1 \\ | \end{pmatrix} \alpha_1 + \cdots \quad \Rightarrow \quad \boxed{\boldsymbol{x} = \boldsymbol{\Psi\alpha}}$$

- Recover (wavelet) coefficients $\boldsymbol{\alpha}$ of image $\boldsymbol{x}$.

- Consider the Laplacian prior on coefficients: $\mathrm{P}(\boldsymbol{\alpha}) \propto \exp\left(-\beta \left\|\boldsymbol{\alpha}\right\|_1\right)$.

- 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

# Compressive sensing

## More sophisticated approaches

- Overcomplete dictionary composed of a concatenation of orthonormal bases:

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

- Constrained vs unconstrained problems.
- Re-weighted versions and the SARA algorithm (Carrillo, McEwen, Wiaux 2012).
- Sparse analysis regularisation problem:

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

Analysis framework

- Analysis problem viewed from a synthesis perspective:

$$\boldsymbol{x}_{\text{analysis}} = \boldsymbol{\Psi} \times \arg\min_{\boldsymbol{\gamma} \in \text{column space}(\boldsymbol{\Psi}^\dagger)} \left[ \left\| \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}^\dagger \boldsymbol{\gamma} \right\|_2^2 + \lambda \left\| \boldsymbol{\gamma} \right\|_1 \right]$$

Analysis as synthesis

# Compressive sensing

## More sophisticated approaches

- Overcomplete dictionary composed of a concatenation of orthonormal bases:

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

- Constrained vs unconstrained problems.
- Re-weighted versions and the SARA algorithm (Carrillo, McEwen, Wiaux 2012).
- Sparse analysis regularisation problem:

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

Analysis framework

- Analysis problem viewed from a synthesis perspective:

$$\boldsymbol{x}_{\text{analysis}} = \boldsymbol{\Psi} \times \arg\min_{\gamma \in \text{column space}(\boldsymbol{\Psi}^\dagger)} \left[ \left\| \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}^\dagger \gamma \right\|_2^2 + \lambda \left\| \gamma \right\|_1 \right]$$

Analysis as synthesis

# Compressive sensing

## More sophisticated approaches

- Overcomplete dictionary composed of a concatenation of orthonormal bases:

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

- Constrained vs unconstrained problems.
- Re-weighted versions and the SARA algorithm (Carrillo, McEwen, Wiaux 2012).
- Sparse analysis regularisation problem:

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

Analysis framework

- Analysis problem viewed from a synthesis perspective:

$$\boldsymbol{x}_{\text{analysis}} = \boldsymbol{\Psi} \times \arg\min_{\boldsymbol{\gamma} \in \text{column space}(\boldsymbol{\Psi}^\dagger)} \left[ \left\| \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}^\dagger \boldsymbol{\gamma} \right\|_2^2 + \lambda \left\| \boldsymbol{\gamma} \right\|_1 \right]$$

Analysis as synthesis

# Compressive sensing

## More sophisticated approaches

- Overcomplete dictionary composed of a concatenation of orthonormal bases:

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

- Constrained vs unconstrained problems.
- Re-weighted versions and the SARA algorithm (Carrillo, McEwen, Wiaux 2012).
- Sparse analysis regularisation problem:

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

Analysis framework

- Analysis problem viewed from a synthesis perspective:

$$\boldsymbol{x}_{\text{analysis}} = \boldsymbol{\Psi} \times \arg\min_{\gamma \in \text{column space}(\boldsymbol{\Psi}^\dagger)} \left[ \left\| \boldsymbol{y} - \boldsymbol{\Phi} \boldsymbol{\Psi}^\dagger \gamma \right\|_2^2 + \lambda \left\| \gamma \right\|_1 \right]$$

Analysis as synthesis

# Compressive sensing

## More sophisticated approaches

- Overcomplete dictionary composed of a concatenation of orthonormal bases:

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

- Constrained vs unconstrained problems.
- Re-weighted versions and the SARA algorithm (Carrillo, McEwen, Wiaux 2012).
- Sparse analysis regularisation problem:

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

Analysis framework

- Analysis problem viewed from a synthesis perspective:

$$\boldsymbol{x}_{\text{analysis}} = \boldsymbol{\Psi} \times \arg\min_{\boldsymbol{\gamma} \in \text{column space}(\boldsymbol{\Psi}^\dagger)} \left[ \left\| \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}^\dagger \boldsymbol{\gamma} \right\|_2^2 + \lambda \left\| \boldsymbol{\gamma} \right\|_1 \right]$$

Analysis as synthesis

# Sampling the full posterior distribution
## Markov Chain Monte Carlo (MCMC)

- Alternative is to sample full posterior distribution $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y})$.
- Provides uncertainly (error) information.
- 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)
- Gibbs sampling applied to radio interferometric imaging (Sutter, Wandelt, McEwen, *et al.* 2014), using methods developed for CMB by Wandelt *et al.* (2005).
  - Assume isotropic Gaussian process prior characterised by power spectrum $C_\ell$.
  - Sample from conditional distributions:

$$\boldsymbol{x}^{i+1} \leftarrow \mathrm{P}(\boldsymbol{x} \mid C_\ell^i, \boldsymbol{y}) \quad \text{and} \quad C_\ell^{i+1} \leftarrow \mathrm{P}(C_\ell \mid \boldsymbol{x}^{i+1}) \, .$$

# Sampling the full posterior distribution
## Markov Chain Monte Carlo (MCMC)

- Alternative is to sample full posterior distribution $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y})$.
- Provides uncertainly (error) information.
- 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)
- Gibbs sampling applied to radio interferometric imaging (Sutter, Wandelt, McEwen, *et al.* 2014), using methods developed for CMB by Wandelt *et al.* (2005).
    - Assume isotropic Gaussian process prior characterised by power spectrum $C_\ell$.
    - Sample from conditional distributions:

$$\boldsymbol{x}^{i+1} \leftarrow \mathrm{P}(\boldsymbol{x} \mid C_\ell^i, \boldsymbol{y}) \quad \text{and} \quad C_\ell^{i+1} \leftarrow \mathrm{P}(C_\ell \mid \boldsymbol{x}^{i+1}) \; .$$

# Sampling the full posterior distribution
## Markov Chain Monte Carlo (MCMC)

- Alternative is to sample full posterior distribution $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y})$.
- Provides uncertainly (error) information.
- 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)
- Gibbs sampling applied to radio interferometric imaging (Sutter, Wandelt, McEwen, *et al.* 2014), using methods developed for CMB by Wandelt *et al.* (2005).
  - Assume isotropic Gaussian process prior characterised by power spectrum $C_\ell$.
  - Sample from conditional distributions:

$$\boldsymbol{x}^{i+1} \leftarrow \mathrm{P}(\boldsymbol{x} \mid C_\ell^i, \boldsymbol{y}) \quad \text{and} \quad C_\ell^{i+1} \leftarrow \mathrm{P}(C_\ell \mid \boldsymbol{x}^{i+1}) \, .$$

# Sampling the full posterior distribution

## Markov Chain Monte Carlo (MCMC)

- Alternative is to sample full posterior distribution $\mathrm{P}(\boldsymbol{x} \,|\, \boldsymbol{y})$.
- Provides uncertainly (error) information.
- 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)
- Gibbs sampling applied to radio interferometric imaging (Sutter, Wandelt, McEwen, *et al.* 2014), using methods developed for CMB by Wandelt *et al.* (2005).
  - Assume isotropic Gaussian process prior characterised by power spectrum $C_\ell$.
  - Sample from conditional distributions:

$$\boldsymbol{x}^{i+1} \leftarrow \mathrm{P}(\boldsymbol{x} \,|\, C_\ell^i, \boldsymbol{y}) \quad \text{and} \quad C_\ell^{i+1} \leftarrow \mathrm{P}(C_\ell \,|\, \boldsymbol{x}^{i+1}) \, .$$

Require MCMC approach to support sparse priors, which shown to be highly effective.

# Outline

# MCMC sampling with gradients
## Langevin dynamics

- Work done by Xiaohao Cai in collaboration with Marcelo Pereyra.
- Consider posteriors of the following form (and more compact notation):

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

- If $g(\boldsymbol{x})$ differentiable can adopt MALA (Langevin dynamics) or HMC (Hamiltonian dynamics) MCMC methods.
- Langevin dynamics model molecular dynamics (includes friction and occasional high velocity collisions that perturb the system).
- Based on Langevin diffusion process $\mathcal{L}(t)$, with $\pi$ 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 sparse priors.

# MCMC sampling with gradients

## Langevin dynamics

- Work done by Xiaohao Cai in collaboration with Marcelo Pereyra.
- Consider posteriors of the following form (and more compact notation):

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \boxed{\pi(\boldsymbol{x})} \propto \exp\left[-\boxed{g(\boldsymbol{x})}\right]$$

Posterior Convex

- If $g(x)$ differentiable can adopt MALA (Langevin dynamics) or HMC (Hamiltonian dynamics) MCMC methods.
- Langevin dynamics model molecular dynamics (includes friction and occasional high velocity collisions that perturb the system).
- Based on Langevin diffusion process $\mathcal{L}(t)$, with $\pi$ 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 sparse priors.

# MCMC sampling with gradients

## Langevin dynamics

- Work done by Xiaohao Cai in collaboration with Marcelo Pereyra.
- Consider posteriors of the following form (and more compact notation):

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

- If $g(\boldsymbol{x})$ differentiable can adopt MALA (Langevin dynamics) or HMC (Hamiltonian dynamics) MCMC methods.
- Langevin dynamics model molecular dynamics (includes friction and occasional high velocity collisions that perturb the system).
- Based on Langevin diffusion process $\mathcal{L}(t)$, with $\pi$ 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 sparse priors.

# MCMC sampling with gradients

## Langevin dynamics

- Work done by Xiaohao Cai in collaboration with Marcelo Pereyra.
- Consider posteriors of the following form (and more compact notation):

$$\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \boxed{\pi(\boldsymbol{x})} \propto \exp\left[-\boxed{g(\boldsymbol{x})}\right]$$

Posterior Convex

- If $g(\boldsymbol{x})$ differentiable can adopt MALA (Langevin dynamics) or HMC (Hamiltonian dynamics) MCMC methods.
- Langevin dynamics model molecular dynamics (includes friction and occasional high velocity collisions that perturb the system).
- Based on Langevin diffusion process $\mathcal{L}(t)$, with $\pi$ 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 sparse priors.

# MCMC sampling with gradients
## Langevin dynamics

- Work done by Xiaohao Cai in collaboration with Marcelo Pereyra.
- Consider posteriors of the following form (and more compact notation):

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

- If $g(\boldsymbol{x})$ differentiable can adopt MALA (Langevin dynamics) or HMC (Hamiltonian dynamics) MCMC methods.
- Langevin dynamics model molecular dynamics (includes friction and occasional high velocity collisions that perturb the system).
- Based on Langevin diffusion process $\mathcal{L}(t)$, with $\pi$ as stationary distribution:

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

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

# Proximity operators

## A brief aside

- Define proximity operator:

$$\mathrm{prox}_g^{\lambda}(\boldsymbol{x}) = \arg\min_{\boldsymbol{u}} \left[ g(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{x}\|^2 / 2\lambda \right]$$

- Generalisation of projection operator:

$$\mathcal{P}_{\mathcal{C}}(\boldsymbol{x}) = \arg\min_{\boldsymbol{u}} \left[ \imath_{\mathcal{C}}(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{x}\|^2 / 2 \right],$$

where $\imath_{\mathcal{C}}(\boldsymbol{u}) = \infty$ if $\boldsymbol{u} \notin \mathcal{C}$ and zero otherwise.

# Proximity operators

## A brief aside

- Define proximity operator:

$$\mathrm{prox}_g^{\lambda}(\boldsymbol{x}) = \arg\min_{\boldsymbol{u}} \left[ g(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{x}\|^2 / 2\lambda \right]$$

- Generalisation of projection operator:

$$\mathcal{P}_{\mathcal{C}}(\boldsymbol{x}) = \arg\min_{\boldsymbol{u}} \left[ \imath_{\mathcal{C}}(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{x}\|^2 / 2 \right],$$

where $\imath_{\mathcal{C}}(\boldsymbol{u}) = \infty$ if $\boldsymbol{u} \notin \mathcal{C}$ and zero otherwise.

# Proximity operators

## A brief aside

- Define proximity operator:

$$\text{prox}_g^\lambda(\boldsymbol{x}) = \arg\min_{\boldsymbol{u}} \left[ g(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{x}\|^2 / 2\lambda \right]$$

- Generalisation of projection operator:

$$\mathcal{P}_\mathcal{C}(\boldsymbol{x}) = \arg\min_{\boldsymbol{u}} \left[ \imath_\mathcal{C}(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{x}\|^2 / 2 \right],$$

where $\imath_\mathcal{C}(\boldsymbol{u}) = \infty$ if $\boldsymbol{u} \notin \mathcal{C}$ and zero otherwise.

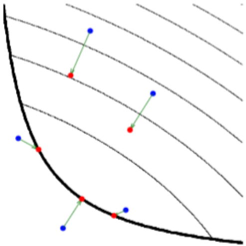


Figure: Illustration of proximity operator [Credit: Parikh & Boyd (2013)]

# Proximal MCMC

## Moreau approximation

- Follow Pereyra (2016a) and consider Moreau approximation of $\pi$:

$$\pi_\lambda(\boldsymbol{x}) = \sup_{\boldsymbol{u}\in\mathbb{R}^N} \pi(\boldsymbol{u}) \exp\Big(-\frac{\|\boldsymbol{u}-\boldsymbol{x}\|^2}{2\lambda}\Big)$$

- Important properties of $\pi_\lambda(\boldsymbol{x})$:

  1. As $\lambda \rightarrow 0$, $\pi_\lambda(\boldsymbol{x}) \rightarrow \pi(\boldsymbol{x})$
  2. $\nabla \log \pi_\lambda(\boldsymbol{x}) = (\text{prox}_g^\lambda(\boldsymbol{x}) - \boldsymbol{x})/\lambda \in \partial \log \pi\big(\text{prox}_g^\lambda(\boldsymbol{x})\big)$

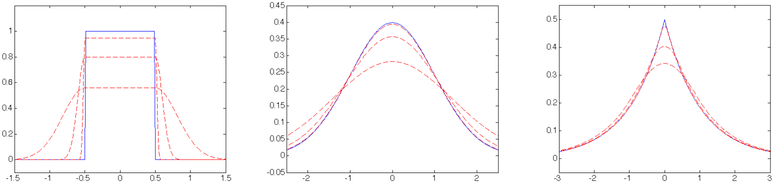


Figure: Illustration of Moreau approximations [Credit: Pereyra (2016a)]

# Proximal MCMC
## Moreau approximation

- Follow Pereyra (2016a) and consider Moreau approximation of $\pi$:

$$\pi_\lambda(\boldsymbol{x}) = \sup_{\boldsymbol{u}\in\mathbb{R}^N} \pi(\boldsymbol{u}) \exp\Big(-\frac{\|\boldsymbol{u}-\boldsymbol{x}\|^2}{2\lambda}\Big)$$

- Important properties of $\pi_\lambda(\boldsymbol{x})$:

1. As $\lambda \to 0, \pi_\lambda(\boldsymbol{x}) \to \pi(\boldsymbol{x})$
2. $\nabla \log \pi_\lambda(\boldsymbol{x}) = (\text{prox}_g^\lambda(\boldsymbol{x}) - \boldsymbol{x})/\lambda \in \partial \log \pi\big(\text{prox}_g^\lambda(\boldsymbol{x})\big)$

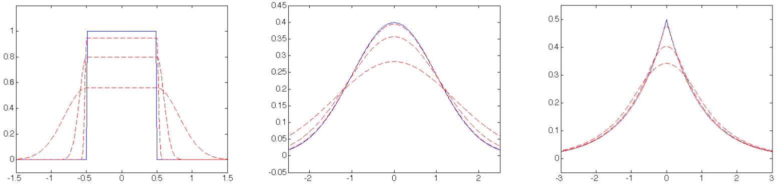



Figure: Illustration of Moreau approximations [Credit: Pereyra (2016a)]

# Proximal MCMC

## P-MALA

- Discretise Langevin differential equation using forward Euler approximation:

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

  where $\delta$ controls the discrete-time increment and $\boldsymbol{w}^{(m)}$ is (unit) Gaussian distributed.

- Apply Moreau approximation and compute gradient by prox:

$$\boldsymbol{l}^{(m+1)} = \operatorname{prox}_g^{\delta/2}(\boldsymbol{l}^{(m)}) + \sqrt{\delta} \boldsymbol{w}^{(m)}.$$

- (After adding a Metropolis-Hastings accept-reject step, get P-MALA with $\boldsymbol{l}^{(m)} \to \pi$.)

- Must compute $\operatorname{prox}_g^{\delta/2}(\cdot)$: develop analytic expression that can be computed rapidly.

# Proximal MCMC

## P-MALA

- Discretise Langevin differential equation using forward Euler approximation:

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

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

where $\delta$ controls the discrete-time increment and $\boldsymbol{w}^{(m)}$ is (unit) Gaussian distributed.

- Apply Moreau approximation and compute gradient by prox:

$$\boldsymbol{l}^{(m+1)} = \mathrm{prox}_g^{\delta/2}(\boldsymbol{l}^{(m)}) + \sqrt{\delta}\boldsymbol{w}^{(m)}.$$

- (After adding a Metropolis-Hastings accept-reject step, get P-MALA with $\boldsymbol{l}^{(m)} \rightarrow \pi$.)

- Must compute $\mathrm{prox}_g^{\delta/2}(\cdot)$: develop analytic expression that can be computed rapidly.

# Proximal MCMC
## P-MALA

- Discretise Langevin differential equation using forward Euler approximation:

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

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

  where $\delta$ controls the discrete-time increment and $\boldsymbol{w}^{(m)}$ is (unit) Gaussian distributed.

- Apply Moreau approximation and compute gradient by prox:

$$\boldsymbol{l}^{(m+1)} = \mathrm{prox}_g^{\delta/2}(\boldsymbol{l}^{(m)}) + \sqrt{\delta}\boldsymbol{w}^{(m)}.$$

- (After adding a Metropolis-Hastings accept-reject step, get P-MALA with $\boldsymbol{l}^{(m)} \rightarrow \pi$.)

- Must compute $\mathrm{prox}_g^{\delta/2}(\cdot)$: develop analytic expression that can be computed rapidly.

# Proximal MCMC
## P-MALA

- Discretise Langevin differential equation using forward Euler approximation:

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

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

where $\delta$ controls the discrete-time increment and $\boldsymbol{w}^{(m)}$ is (unit) Gaussian distributed.

- Apply Moreau approximation and compute gradient by prox:

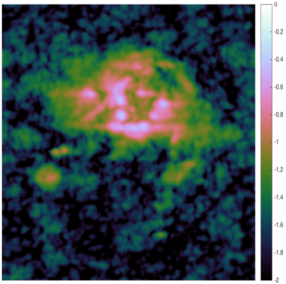
$$\boldsymbol{l}^{(m+1)} = \mathrm{prox}_g^{\delta/2}(\boldsymbol{l}^{(m)}) + \sqrt{\delta}\boldsymbol{w}^{(m)}.$$

- (After adding a Metropolis-Hastings accept-reject step, get P-MALA with $\boldsymbol{l}^{(m)} \rightarrow \pi$.)

- Must compute $\mathrm{prox}_g^{\delta/2}(\cdot)$: develop analytic expression that can be computed rapidly.

# Proximal MCMC

## Preliminary results on simulations

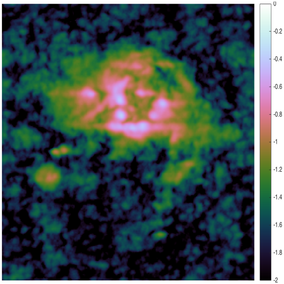



(a) Dirty image

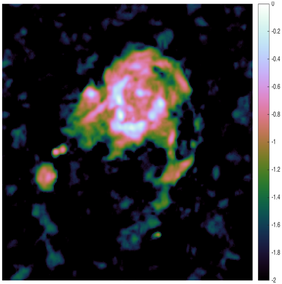
Figure: HII region of M31

# Proximal MCMC

## Preliminary results on simulations




(a) Dirty image

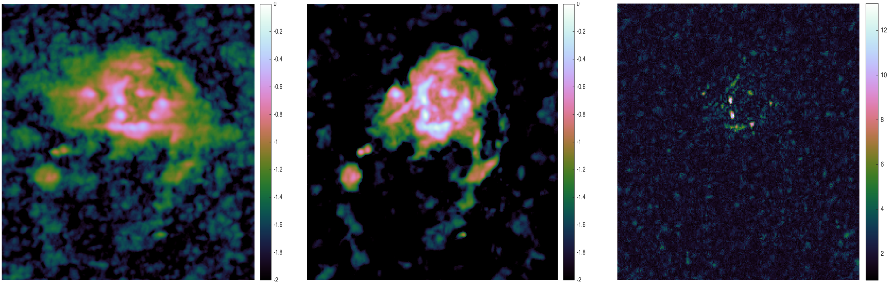



(b) Mean recovered image

Figure: HII region of M31

# Proximal MCMC

## Preliminary results on simulations

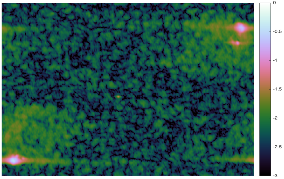



(a) Dirty image (b) Mean recovered image (c) Standard deviation image

Figure: HII region of M31

# Proximal MCMC

## Preliminary results on simulations

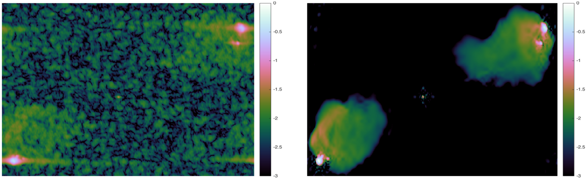



(a) Dirty image

Figure: Cygnus A

# Proximal MCMC

## Preliminary results on simulations



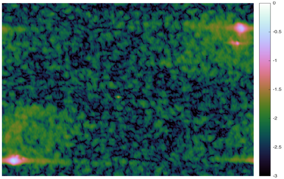

(a) Dirty image

(b) Mean recovered image

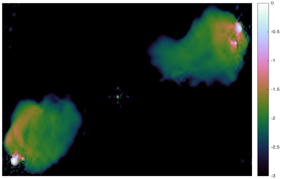
Figure: Cygnus A

# Proximal MCMC

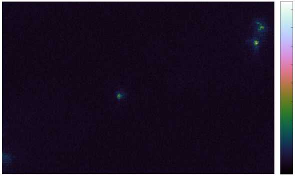
## Preliminary results on simulations




(a) Dirty image




(b) Mean recovered image

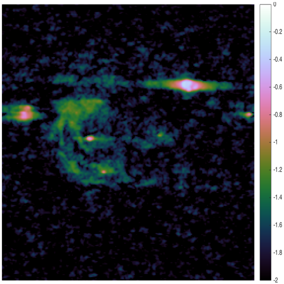


(c) Standard deviation image

Figure: Cygnus A

# Proximal MCMC

## Preliminary results on simulations

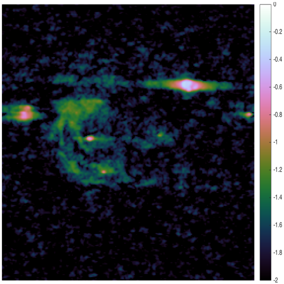



(a) Dirty image

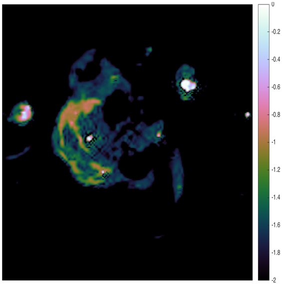
Figure: Supernova remnant W28

# Proximal MCMC

## Preliminary results on simulations




(a) Dirty image

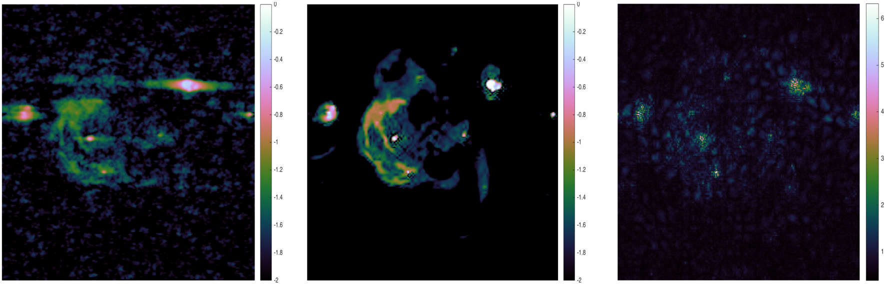



(b) Mean recovered image

Figure: Supernova remnant W28

# Proximal MCMC

## Preliminary results on simulations

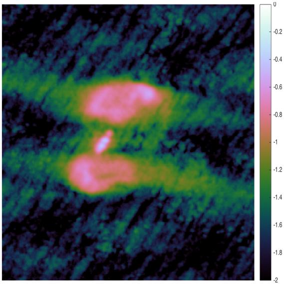



(a) Dirty image (b) Mean recovered image (c) Standard deviation image

Figure: Supernova remnant W28

# Proximal MCMC

## Preliminary results on simulations

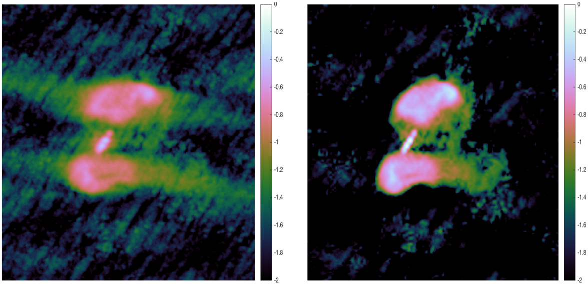



(a) Dirty image

Figure: 3C288

# Proximal MCMC

## Preliminary results on simulations

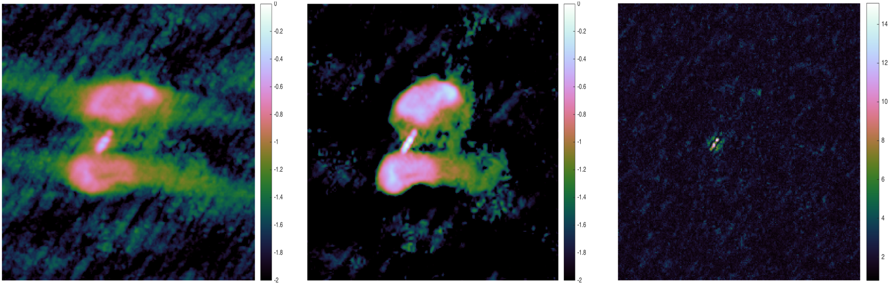



(a) Dirty image (b) Mean recovered image

Figure: 3C288

# Proximal MCMC

## Preliminary results on simulations




(a) Dirty image (b) Mean recovered image (c) Standard deviation image

Figure: 3C288

# Outline

1. Bayesian interpretations of interferometric imaging techniques
2. Proximal MCMC sampling
3. Sparse regularisation with Bayesian credible regions

# Bayesian credible regions

## Computation

- Combine error estimation with fast sparse regularisation (*cf.* compressive sensing) methods.
- Let $C_\alpha$ denote a Bayesian credible region with confidence level $(1-\alpha)\%$:

$$\mathrm{P}(\boldsymbol{x} \in C_\alpha \,|\, \boldsymbol{y}) = \int_{\boldsymbol{x} \in C_\alpha} p(\boldsymbol{x}|\boldsymbol{y}) \,\mathrm{d}\boldsymbol{x} = 1 - \alpha$$

- Define $C_\alpha$ by posterior iso-contour:

$$C_\alpha := \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$$

- Analytic approximation of $\gamma_\alpha$ derived in Pereyra (2016b):

$$\tilde{\gamma}_\alpha = g_{\boldsymbol{y}}(\boldsymbol{x}_{\mathrm{map}}) + N(\tau_\alpha + 1),$$

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

- Compute $\boldsymbol{x}_{\mathrm{map}}$ by sparse regularisation and estimate Bayesian credible regions.

# Bayesian credible regions

## Computation

- Combine error estimation with fast sparse regularisation (*cf.* compressive sensing) methods.
- Let $C_\alpha$ denote a Bayesian credible region with confidence level $(1-\alpha)\%$:

$$\mathrm{P}(\boldsymbol{x} \in C_\alpha \,|\, \boldsymbol{y}) = \int_{\boldsymbol{x} \in C_\alpha} p(\boldsymbol{x}|\boldsymbol{y}) \, \mathrm{d}\boldsymbol{x} = 1 - \alpha$$

- Define $C_\alpha$ by posterior iso-contour:

$$C_\alpha := \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$$

- Analytic approximation of $\gamma_\alpha$ derived in Pereyra (2016b):

$$\tilde{\gamma}_\alpha = g_{\boldsymbol{y}}(\boldsymbol{x}_{\mathrm{map}}) + N(\tau_\alpha + 1) \,,$$

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

- Compute $\boldsymbol{x}_{\mathrm{map}}$ by sparse regularisation and estimate Bayesian credible regions.

# Bayesian credible regions

## Computation

- Combine error estimation with fast sparse regularisation (*cf.* compressive sensing) methods.
- Let $C_\alpha$ denote a Bayesian credible region with confidence level $(1-\alpha)\%$:

$$\mathrm{P}(\boldsymbol{x} \in C_\alpha \,|\, \boldsymbol{y}) = \int_{\boldsymbol{x} \in C_\alpha} p(\boldsymbol{x}|\boldsymbol{y}) \, \mathrm{d}\boldsymbol{x} = 1 - \alpha$$

- Define $C_\alpha$ by posterior iso-contour:

$$C_\alpha := \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$$

- Analytic approximation of $\gamma_\alpha$ derived in Pereyra (2016b):

$$\tilde{\gamma}_\alpha = g_{\boldsymbol{y}}(\boldsymbol{x}_{\mathrm{map}}) + N(\tau_\alpha + 1) ,$$

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

- Compute $\boldsymbol{x}_{\mathrm{map}}$ by sparse regularisation and estimate Bayesian credible regions.

# Bayesian credible regions

## Computation

- Combine error estimation with fast sparse regularisation (*cf.* compressive sensing) methods.
- Let $C_\alpha$ denote a Bayesian credible region with confidence level $(1-\alpha)\%$:

$$\mathrm{P}(\boldsymbol{x} \in C_\alpha \,|\, \boldsymbol{y}) = \int_{\boldsymbol{x} \in C_\alpha} p(\boldsymbol{x}|\boldsymbol{y}) \, \mathrm{d}\boldsymbol{x} = 1 - \alpha$$

- Define $C_\alpha$ by posterior iso-contour:

$$\boxed{C_\alpha := \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}}$$

- Analytic approximation of $\gamma_\alpha$ derived in Pereyra (2016b):

$$\boxed{\tilde{\gamma}_\alpha = g_{\boldsymbol{y}}(\boldsymbol{x}_{\mathrm{map}}) + N(\tau_\alpha + 1)} ,$$

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

- Compute $\boldsymbol{x}_{\mathrm{map}}$ by sparse regularisation and estimate Bayesian credible regions.

# Bayesian credible regions

## Computation

- Combine error estimation with fast sparse regularisation (*cf.* compressive sensing) methods.
- Let $C_\alpha$ denote a Bayesian credible region with confidence level $(1-\alpha)\%$:

$$\mathrm{P}(\boldsymbol{x} \in C_\alpha \,|\, \boldsymbol{y}) = \int_{\boldsymbol{x} \in C_\alpha} p(\boldsymbol{x}|\boldsymbol{y}) \, \mathrm{d}\boldsymbol{x} = 1 - \alpha$$

- Define $C_\alpha$ by posterior iso-contour:

$$C_\alpha := \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$$

- Analytic approximation of $\gamma_\alpha$ derived in Pereyra (2016b):

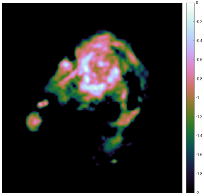
$$\tilde{\gamma}_\alpha = g_{\boldsymbol{y}}(\boldsymbol{x}_{\mathrm{map}}) + N(\tau_\alpha + 1) ,$$

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

- Compute $\boldsymbol{x}_{\mathrm{map}}$ by sparse regularisation and estimate Bayesian credible regions.

# Bayesian credible regions

## Preliminary results on simulations

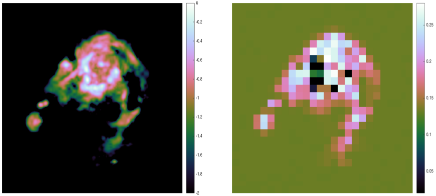



(a) Recovered image

Figure: HII region of M31

# Bayesian credible regions

## Preliminary results on simulations



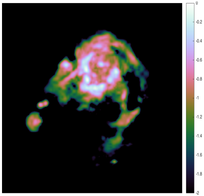
(a) Recovered image

(b) Credible intervals for regions of size $10 \times 10$

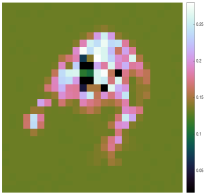
Figure: HII region of M31

# Bayesian credible regions

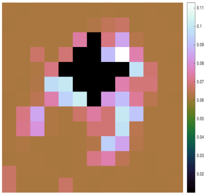
## Preliminary results on simulations



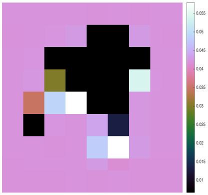
(a) Recovered image



(b) Credible intervals for regions of size $10 \times 10$



(c) Credible intervals for regions of size $20 \times 20$

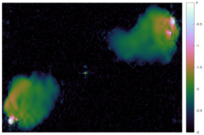


(d) Credible intervals for regions of size $30 \times 30$

Figure: HII region of M31

# Bayesian credible regions

## Preliminary results on simulations

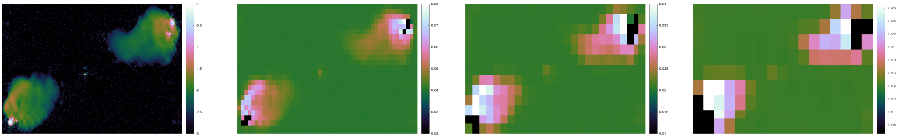


(a) Recovered image

Figure: Cygnus A

# Bayesian credible regions

## Preliminary results on simulations



(a) Recovered image

(b) Credible intervals for regions of size $10 \times 10$

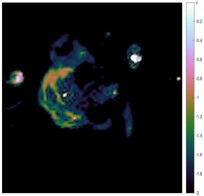
(c) Credible intervals for regions of size $20 \times 20$

(d) Credible intervals for regions of size $30 \times 30$

Figure: Cygnus A

# Bayesian credible regions

## Preliminary results on simulations

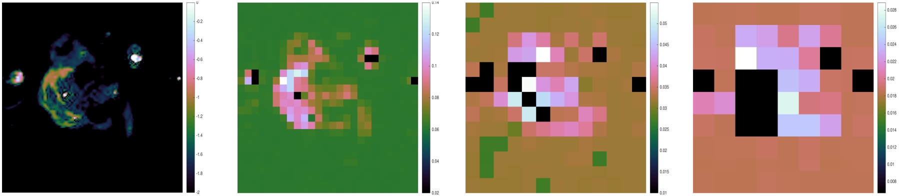



(a) Recovered image

Figure: Supernova remnant W28

# Bayesian credible regions

## Preliminary results on simulations



(a) Recovered image

(b) Credible intervals for regions of size $10 \times 10$

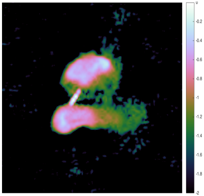
(c) Credible intervals for regions of size $20 \times 20$

(d) Credible intervals for regions of size $30 \times 30$

Figure: Supernova remnant W28

# Bayesian credible regions

## Preliminary results on simulations

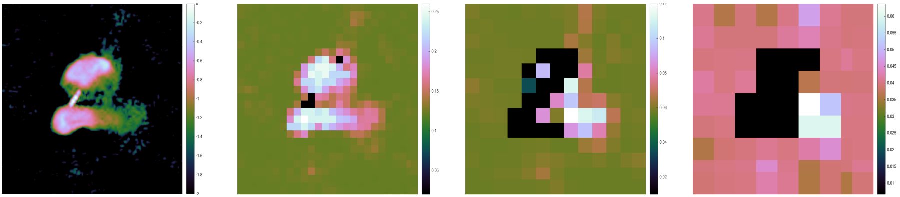



(a) Recovered image

Figure: 3C288

# Bayesian credible regions

## Preliminary results on simulations



(a) Recovered image

(b) Credible intervals for regions of size $10 \times 10$

(c) Credible intervals for regions of size $20 \times 20$

(d) Credible intervals for regions of size $30 \times 30$

Figure: 3C288

# Hypothesis testing

## Method

- Is structure in an image physical or an artefact?
- Can we make precise statistical statements?
- Perform hypothesis tests using Bayesian credible regions (Pereyra 2016b).

# Hypothesis testing

## Method

- Is structure in an image physical or an artefact?
- Can we make precise statistical statements?
- Perform hypothesis tests using Bayesian credible regions (Pereyra 2016b).

# Hypothesis testing
## Method

- Is structure in an image physical or an artefact?
- Can we make precise statistical statements?
- Perform hypothesis tests using Bayesian credible regions (Pereyra 2016b).

# Hypothesis testing
## Method

- Is structure in an image physical or an artefact?
- Can we make precise statistical statements?
- Perform hypothesis tests using Bayesian credible regions (Pereyra 2016b).

### Hypothesis testing of physical structure

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

# Hypothesis testing
## Method

- Is structure in an image physical or an artefact?
- Can we make precise statistical statements?
- Perform hypothesis tests using Bayesian credible regions (Pereyra 2016b).

### Hypothesis testing of physical structure

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

# Hypothesis testing

## Method

- Is structure in an image physical or an artefact?
- Can we make precise statistical statements?
- Perform hypothesis tests using Bayesian credible regions (Pereyra 2016b).

### Hypothesis testing of physical structure

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

# Hypothesis testing

## Method

- Is structure in an image physical or an artefact?
- Can we make precise statistical statements?
- Perform hypothesis tests using Bayesian credible regions (Pereyra 2016b).

### Hypothesis testing of physical structure

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

# Hypothesis testing
## Method

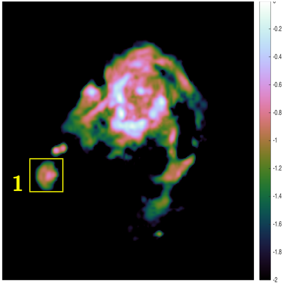
- Is structure in an image physical or an artefact?
- Can we make precise statistical statements?
- Perform hypothesis tests using Bayesian credible regions (Pereyra 2016b).

### Hypothesis testing of physical structure

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

# Hypothesis testing

## Preliminary results on simulations

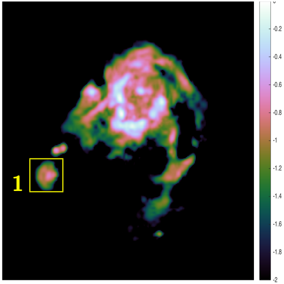



(a) Recovered image

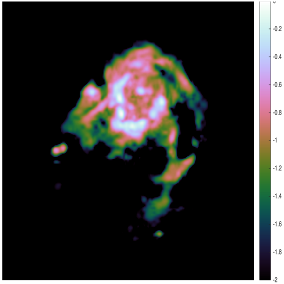
Figure: HII region of M31

# Hypothesis testing

## Preliminary results on simulations




(a) Recovered image

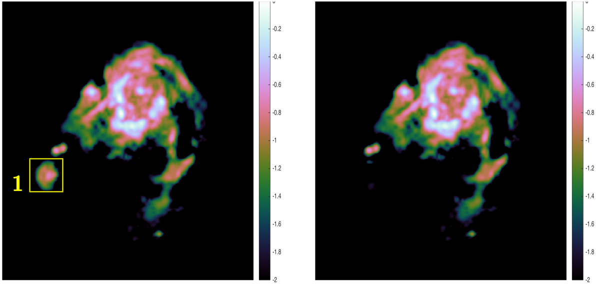



(b) Surrogate with region removed

Figure: HII region of M31

# Hypothesis testing

## Preliminary results on simulations




(a) Recovered image

(b) Surrogate with region removed

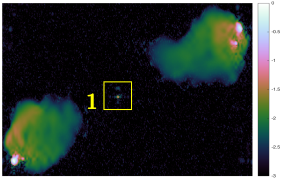
Figure: HII region of M31

Reject null hypothesis

⇒ structure physical

# Hypothesis testing

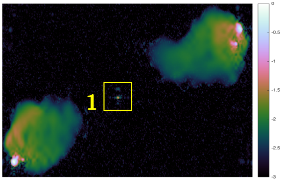
## Preliminary results on simulations



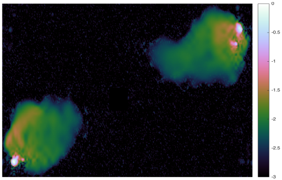

(a) Recovered image

Figure: Cygnus A

# Hypothesis testing
## Preliminary results on simulations




(a) Recovered image

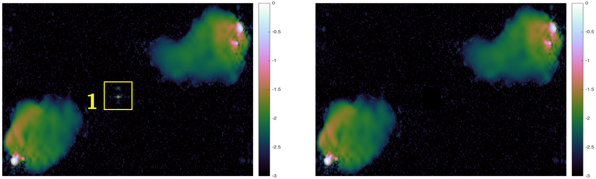



(b) Surrogate with region removed

Figure: Cygnus A

# Hypothesis testing

## Preliminary results on simulations




(a) Recovered image

(b) Surrogate with region removed

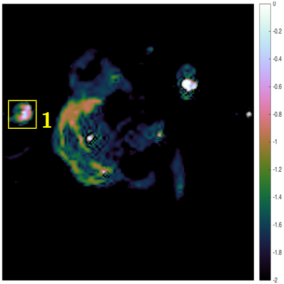
Figure: Cygnus A

Cannot reject null hypothesis

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

# Hypothesis testing

## Preliminary results on simulations

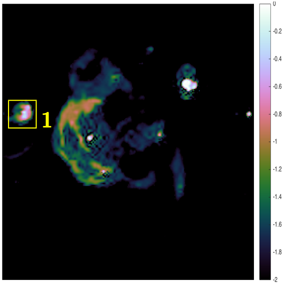



(a) Recovered image

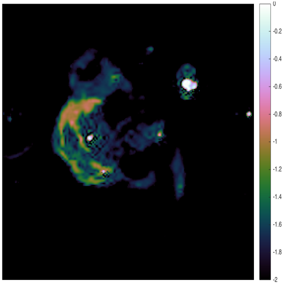
Figure: Supernova remnant W28

# Hypothesis testing

## Preliminary results on simulations




(a) Recovered image

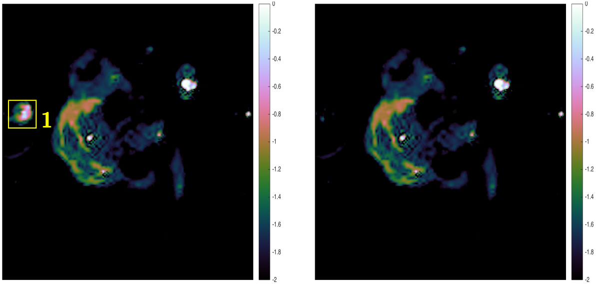



(b) Surrogate with region removed

Figure: Supernova remnant W28

# Hypothesis testing

## Preliminary results on simulations




(a) Recovered image

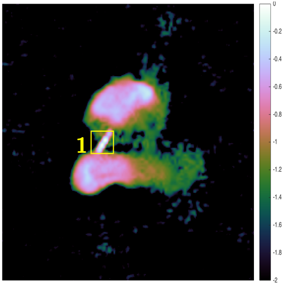
(b) Surrogate with region removed

Figure: Supernova remnant W28

Reject null hypothesis

$\Rightarrow$ structure physical

# Hypothesis testing
## Preliminary results on simulations

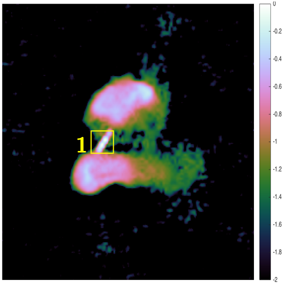



(a) Recovered image

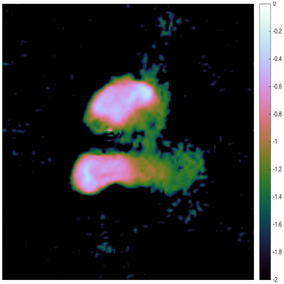
Figure: 3C288

# Hypothesis testing

## Preliminary results on simulations




(a) Recovered image

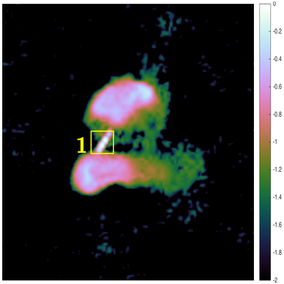



(b) Surrogate with region removed

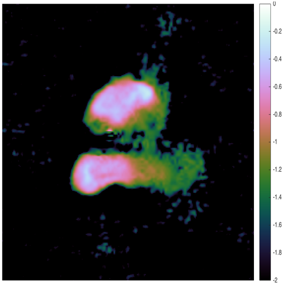
Figure: 3C288

# Hypothesis testing

## Preliminary results on simulations




(a) Recovered image




(b) Surrogate with region removed

Figure: 3C288

Reject null hypothesis

$\Rightarrow$ structure physical

# Conclusions

- Sparse priors (*cf.* compressive sensing) shown to be highly effective (see talks by Luke Pratley, Alex Onose, Vijay Kartik).
- Also seek statistical interpretation to recover error information.
- Proximal MCMC sampling can support sparse priors in full statistical framework (P-MALA).
- Combine error estimation with fast sparse regularisation:
  - Recover Bayesian credible regions.
  - Perform hypothesis testing to test whether structure physical.

# Conclusions

- Sparse priors (*cf.* compressive sensing) shown to be highly effective (see talks by Luke Pratley, Alex Onose, Vijay Kartik).
- Also seek statistical interpretation to recover error information.
- Proximal MCMC sampling can support sparse priors in full statistical framework (P-MALA).
- Combine error estimation with fast sparse regularisation:
  - Recover Bayesian credible regions.
  - Perform hypothesis testing to test whether structure physical.

# Conclusions

- Sparse priors (*cf.* compressive sensing) shown to be highly effective (see talks by Luke Pratley, Alex Onose, Vijay Kartik).
- Also seek statistical interpretation to recover error information.
- Proximal MCMC sampling can support sparse priors in full statistical framework (P-MALA).
- Combine error estimation with fast sparse regularisation:
  - Recover Bayesian credible regions.
  - Perform hypothesis testing to test whether structure physical.

# Conclusions

- Sparse priors (*cf.* compressive sensing) shown to be highly effective (see talks by Luke Pratley, Alex Onose, Vijay Kartik).
- Also seek statistical interpretation to recover error information.
- Proximal MCMC sampling can support sparse priors in full statistical framework (P-MALA).
- Combine error estimation with fast sparse regularisation:
  - Recover Bayesian credible regions.
  - Perform hypothesis testing to test whether structure physical.