Next-generation radio interferometric imaging for the SKA era

From Bayesian inference and compressed sensing, to big-data, to uncertainty quantification

Jason McEwen
www.jasonmcewen.org
@jasonmcewen

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

University of Manchester, March 2017

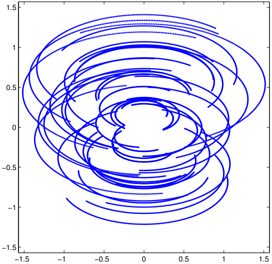
Outline

1. A unified framework for radio interferometric imaging
2. Compressive sensing for SKA imaging
3. 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:

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

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.

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.

Radio interferometric inverse problem

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

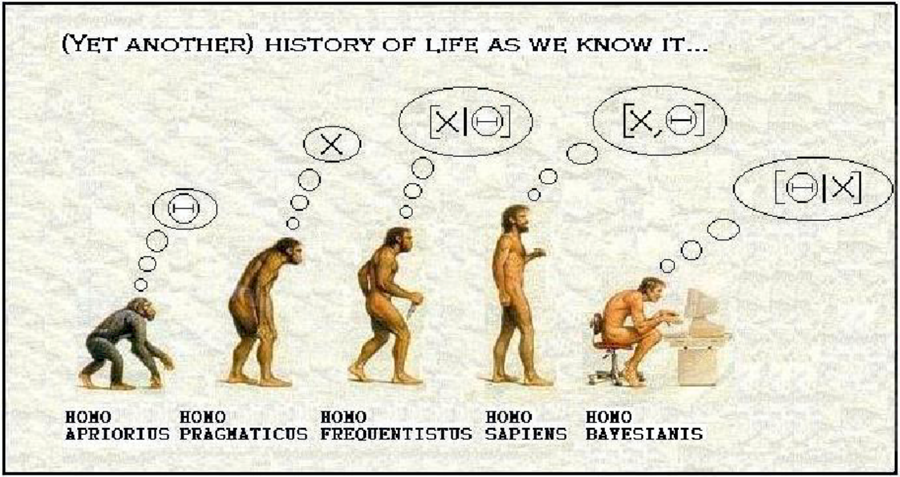
$$\boxed{\boldsymbol{y} = \boldsymbol{\Phi} \boldsymbol{x} + \boldsymbol{n}},$$

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.

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:

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

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:

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

Bayes Theorem

- Bayes theorem in words:

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

- How do we perform Bayesian inference in practice?

  ⇒ 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)

Bayes in practice

MAP and MCMC sampling

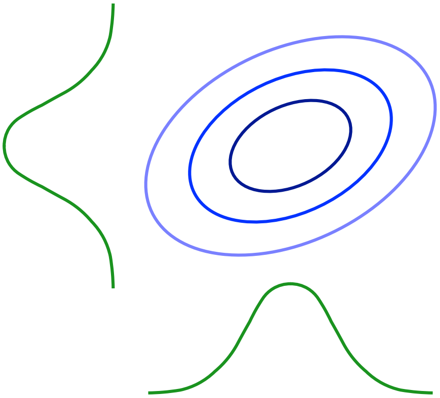


Figure: Probability distribution to explore in 2D

Bayes in practice

MAP and MCMC sampling

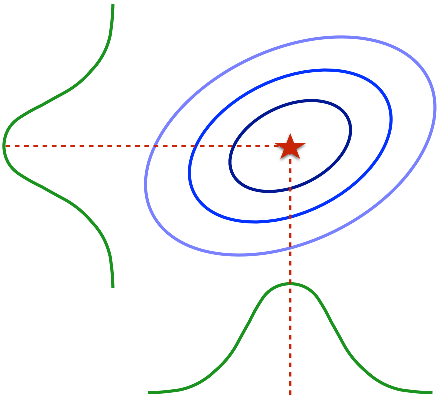


Figure: Maximum a-posteriori (MAP) estimate

Bayes in practice

MAP and MCMC sampling

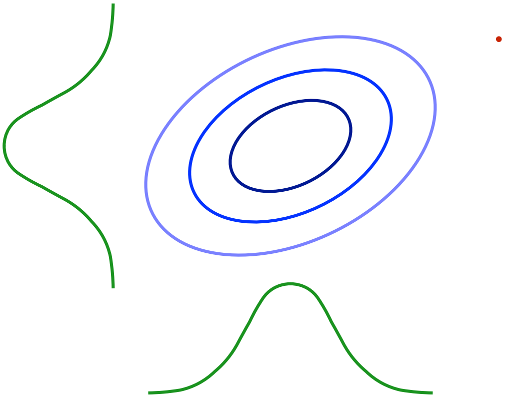


Figure: Markov Chain Monte Carlo (MCMC) sampling

Bayes in practice
MAP and MCMC sampling

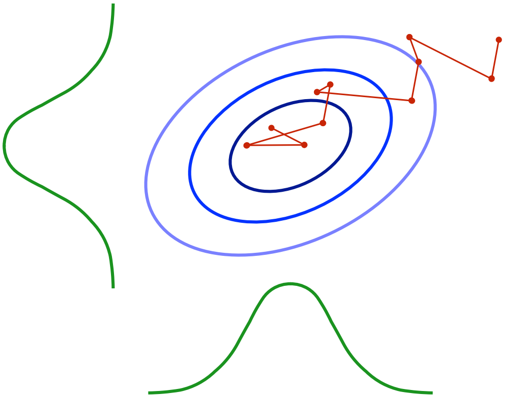


Figure: Markov Chain Monte Carlo (MCMC) sampling

Bayes in practice
MAP and MCMC sampling

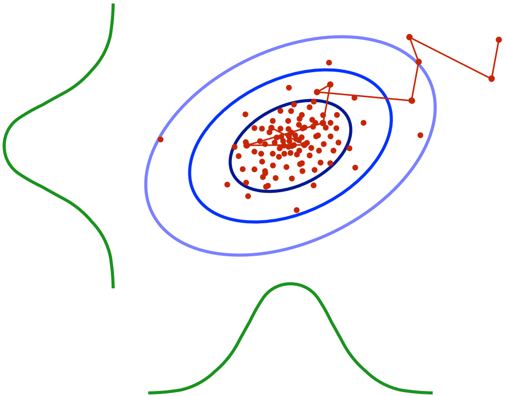


Figure: Markov Chain Monte Carlo (MCMC) sampling

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):

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

- 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:

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

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:

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

Likelihood

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

Prior

- 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:

MAP estimation and regularisation

Hint: they're the same thing!

- Start with Bayes Theorem (ignore normalising evidence):

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

Define likelihood (assuming Gaussian noise) and prior:

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

Likelihood

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

Prior

Consider log-posterior:

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

- MAP estimator:

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

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:

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

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:

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

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:

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

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:

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

Norms often considered for regularisation

- Recall norms given by:

$$\|\boldsymbol{\alpha}\|_2^2 = \sum_i |\boldsymbol{\alpha}_i|^2 \qquad \|\boldsymbol{\alpha}\|_1 = \sum_i |\boldsymbol{\alpha}_i| \qquad \|\boldsymbol{\alpha}\|_0 = \text{no. non-zero elements}$$

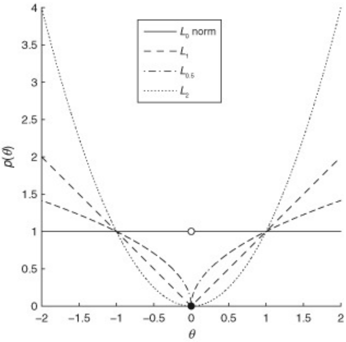


Figure: Norms in 1D **[Credit: Qiao 2014]**

Norms often considered for regularisation

- Recall norms given by:

$$\|\boldsymbol{\alpha}\|_2^2 = \sum_i |\boldsymbol{\alpha}_i|^2 \qquad \|\boldsymbol{\alpha}\|_1 = \sum_i |\boldsymbol{\alpha}_i| \qquad \|\boldsymbol{\alpha}\|_0 = \text{no. non-zero elements}$$

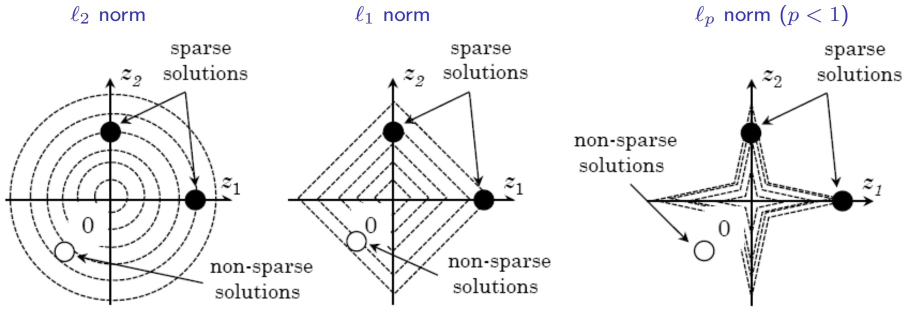


Figure: Norms in 2D **[Credit: Kudo *et al.* 2013]**

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:

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

- MEM

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

  Corresponding MAP estimator is:

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

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

  Corresponding MAP estimator is:

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

- MEM

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

  Corresponding MAP estimator is:

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

(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:

$$\boldsymbol{x}_{\mathrm{clean}} \simeq \underset{\boldsymbol{x}}{\arg\min}\left[\left\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\right\|_2^2 + \lambda \left\|\boldsymbol{x}\right\|_0\right]$$

- MEM

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

  Corresponding MAP estimator is:

$$\boldsymbol{x}_{\mathrm{mem}} \simeq \underset{\boldsymbol{x}}{\arg\min}\left[\left\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\right\|_2^2 + \lambda\, \boldsymbol{x}^\dagger \log \boldsymbol{x}\right]$$

(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:

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

- MEM

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

  Corresponding MAP estimator is:

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

(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: $\text{P}(\boldsymbol{x}) \propto \exp\left(-\beta \, \|\boldsymbol{x}\|_0\right)$.

  Corresponding MAP estimator is:

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

- MEM

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

  Corresponding MAP estimator is:

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

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

Compressive sensing as MAP estimator

- Naive compressive sensing

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

  Corresponding MAP estimator is:

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

(This is one possible Bayesian interpretation of compressive sensing but there are others.)

Compressive sensing as MAP estimator

- Naive compressive sensing

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

  Corresponding MAP estimator is:

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

(This is one possible Bayesian interpretation of compressive sensing but there are others.)

Compressive sensing as MAP estimator

- Naive compressive sensing

  Consider the Laplacian prior: $\text{P}(\boldsymbol{x}) \propto \exp\left(-\beta \left\|\boldsymbol{x}\right\|_1\right)$.

  Corresponding MAP estimator is:

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

(This is one possible Bayesian interpretation of compressive sensing but there are others.)

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 α of image 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):

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

- 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):

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

- 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:

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

Synthesis framework

Compressive sensing

Analysis framework

- 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):

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

Analysis framework

(For orthogonal bases $\Omega = \Psi^\dagger$ and the two approaches are identical.)

Compressive sensing

Analysis framework

- 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):

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

Analysis framework

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

Compressive sensing

Analysis vs synthesis

- Synthesis-based approach is more general, while analysis-based approach more restrictive.

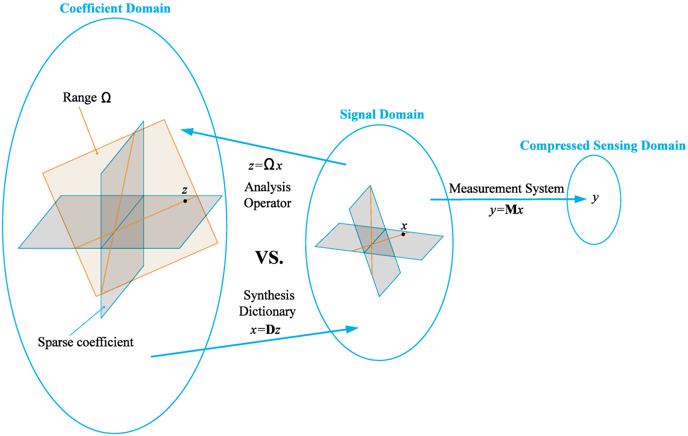


Figure: Analysis- and synthesis-based approaches **[Credit: Nam *et al.* (2012)]**

Compressive sensing

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:

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

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:

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

SARA

Compressive sensing
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:

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

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:

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

SARA

Compressive sensing
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:

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

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:

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

SARA

Outline

Public open-source codes

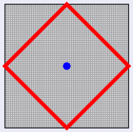
PURIFY code http://basp-group.github.io/purify/

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.

Robust application of PURIFY to real interferometric observations

- Robust sparse image reconstruction of radio interferometric observations with PURIFY (Pratley, McEwen, *et al.* 2016; arXiv:1610.02400).
- All parameters are set automatically (but can be refined).

Table: Description of main user parameters for using PURIFY to reconstruct an observation.

Parameter	PURIFY option	Description	Value
η	-l2_bound	Parameterisation of the fidelity constraint: $\epsilon_\eta = \eta\sqrt{M}\sigma_n$.	$\eta = 1.4$ (default); $\eta \in [1, 10]$ (typical).
β	-beta	Parameterisation of the step size of the algorithm: $\tilde{\gamma}_i = \beta\|\Psi^\dagger x^{(i)}\|_{\ell_\infty}$ (default). One can also fix $\gamma = \beta\|\Psi^\dagger x^{(0)}\|_{\ell_\infty}$.	$\beta = 10^{-3}$ (default)
$\delta_{\rm adapt}$	-relative_gamma_adapt	Relative difference criteria for adapting γ_i.	$\delta_{\rm adapt} = 0.01$ (default).
$i_{\rm adapt}$	-adapt_iter	Number of iterations to consider adapting the step size γ_i (should be before convergence).	$i_{\rm adapt} = 100$ (default).
δ	-relative_variation	Relative difference convergence criteria on the ℓ_2-norm of the solution: $\frac{\|x^{(i)} - x^{(i-1)}\|_{\ell_2}}{\|x^{(i)}\|_{\ell_2}} \leq \delta$.	$\delta = 5 \times 10^{-3}$ (default).
ξ	-residual_convergence	Convergence criteria on the ℓ_2 residual norm: $\|y - \Phi x\|_{\ell_2} \leq \xi\epsilon_\eta$	$\xi = 1$ (default); require $\xi \geq 1$.
$i_{\rm max}$	-niters	Maximum number of iterations.	$i_{\rm max} = \infty$ (default).

Robust application of PURIFY to real interferometric observations

- Robust sparse image reconstruction of radio interferometric observations with PURIFY (Pratley, McEwen, *et al.* 2016; arXiv:1610.02400).
- All parameters are set automatically (but can be refined).

Table: Description of main user parameters for using PURIFY to reconstruct an observation.

Parameter	PURIFY option	Description	Value
η	-l2_bound	**Parameterisation of the fidelity constraint:** $\epsilon_\eta = \eta\sqrt{M}\sigma_n$.	$\eta = 1.4$ **(default)**; $\eta \in [1, 10]$ **(typical)**.
β	-beta	**Parameterisation of the step size of the algorithm:** $\tilde{\gamma}_i = \beta \Vert \Psi^\dagger x^{(i)} \Vert_{\ell_\infty}$ **(default). One can also fix** $\gamma = \beta \Vert \Psi^\dagger x^{(0)} \Vert_{\ell_\infty}$.	$\beta = 10^{-3}$ **(default)**
δ_{adapt}	-relative_gamma_adapt	**Relative difference criteria for adapting** γ_i.	$\delta_{\text{adapt}} = 0.01$ **(default)**.
i_{adapt}	-adapt_iter	**Number of iterations to consider adapting the step size** γ_i **(should be before convergence).**	$i_{\text{adapt}} = 100$ **(default)**.
δ	-relative_variation	**Relative difference convergence criteria on the** ℓ_2**-norm of the solution:** $\frac{\Vert x^{(i)} - x^{(i-1)} \Vert_{\ell_2}}{\Vert x^{(i)} \Vert_{\ell_2}} \leq \delta$.	$\delta = 5 \times 10^{-3}$ **(default)**.
ξ	-residual_convergence	**Convergence criteria on the** ℓ_2 **residual norm:** $\Vert y - \Phi x \Vert_{\ell_2} \leq \xi \epsilon_\eta$	$\xi = 1$ **(default); require** $\xi \geq 1$.
i_{max}	-niters	**Maximum number of iterations.**	$i_{\text{max}} = \infty$ **(default)**.

Robust application of PURIFY to real interferometric observations

- Robust sparse image reconstruction of radio interferometric observations with PURIFY (Pratley, McEwen, *et al.* 2016; arXiv:1610.02400).
- All parameters are set automatically (but can be refined).

Table: Description of main user parameters for using PURIFY to reconstruct an observation.

Parameter	PURIFY option	Description	Value
η	-l2_bound	**Parameterisation of the fidelity constraint:** $\epsilon_\eta = \eta\sqrt{M}\sigma_n$.	$\eta = 1.4$ **(default)**; $\eta \in [1, 10]$ **(typical)**.
β	-beta	**Parameterisation of the step size of the algorithm:** $\tilde{\gamma}_i = \beta\Vert\boldsymbol{\Psi}^\dagger \boldsymbol{x}^{(i)}\Vert_{\ell_\infty}$ **(default). One can also fix** $\gamma = \beta\Vert\boldsymbol{\Psi}^\dagger \boldsymbol{x}^{(0)}\Vert_{\ell_\infty}$.	$\beta = 10^{-3}$ **(default)**
$\delta_{\rm adapt}$	-relative_gamma_adapt	**Relative difference criteria for adapting** γ_i.	$\delta_{\rm adapt} = 0.01$ **(default)**.
$i_{\rm adapt}$	-adapt_iter	**Number of iterations to consider adapting the step size** γ_i **(should be before convergence).**	$i_{\rm adapt} = 100$ **(default)**.
δ	-relative_variation	**Relative difference convergence criteria on the** ℓ_2**-norm of the solution:** $\frac{\Vert\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i-1)}\Vert_{\ell_2}}{\Vert\boldsymbol{x}^{(i)}\Vert_{\ell_2}} \leq \delta$.	$\delta = 5 \times 10^{-3}$ **(default)**.
ξ	-residual_convergence	**Convergence criteria on the** ℓ_2 **residual norm:** $\Vert\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\Vert_{\ell_2} \leq \xi\epsilon_\eta$	$\xi = 1$ **(default); require** $\xi \geq 1$.
$i_{\rm max}$	-niters	**Maximum number of iterations.**	$i_{\rm max} = \infty$ **(default)**.

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

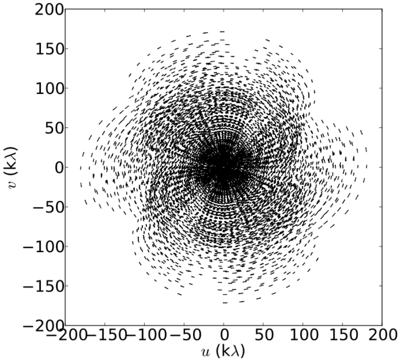
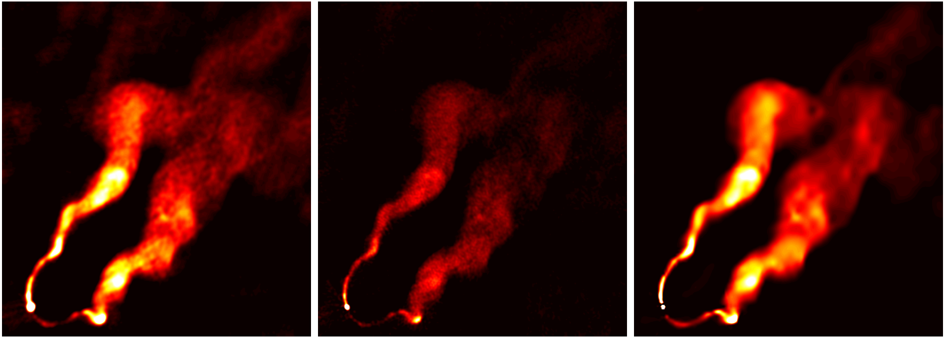


Figure: VLA visibility coverage for 3C129

PURIFY reconstruction
VLA observation of 3C129

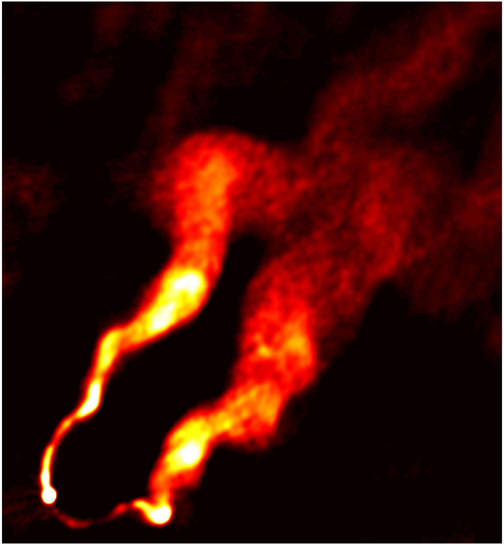


(a) CLEAN (natural) (b) CLEAN (uniform) (c) PURIFY

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

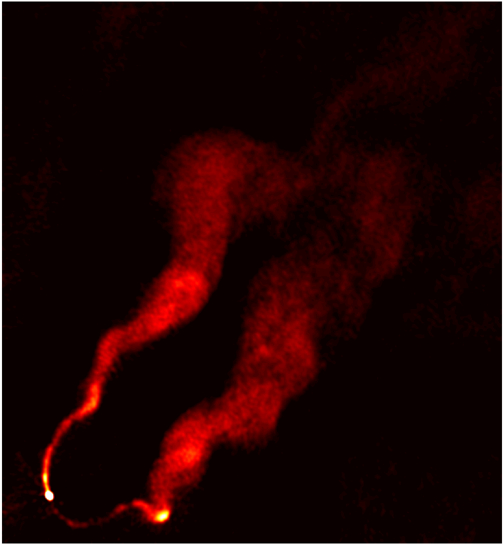
PURIFY reconstruction

VLA observation of 3C129 imaged by CLEAN (natural)



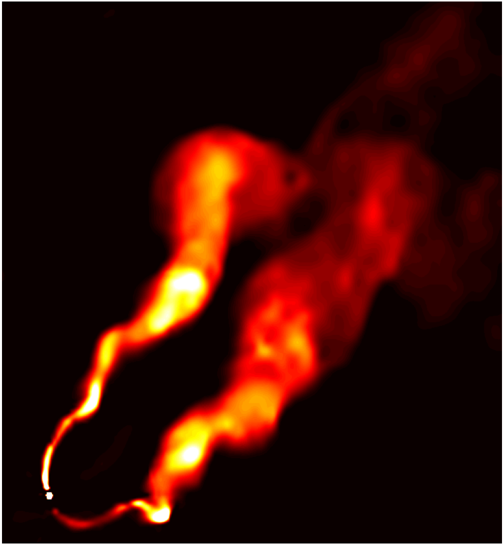
PURIFY reconstruction

VLA observation of 3C129 images by CLEAN (uniform)



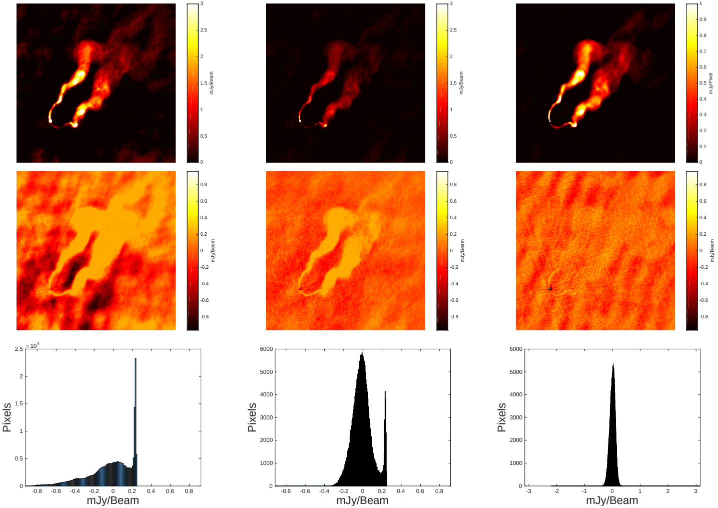
PURIFY reconstruction

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

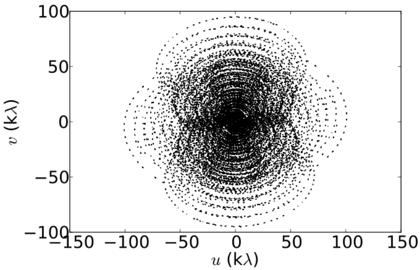
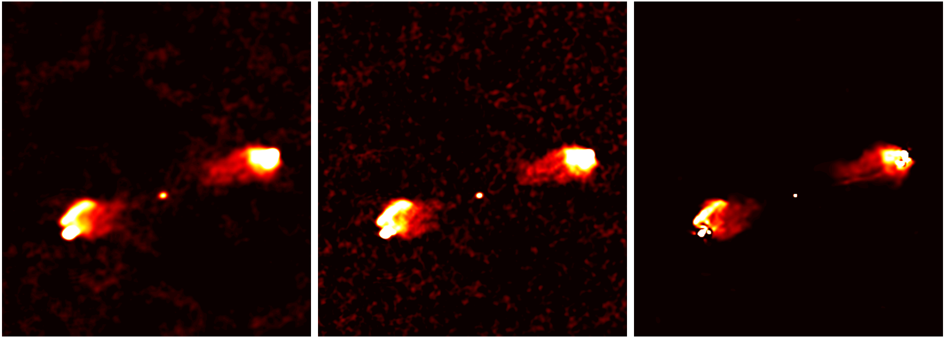


Figure: VLA visibility coverage for Cygnus A

PURIFY reconstruction

VLA observation of Cygnus A

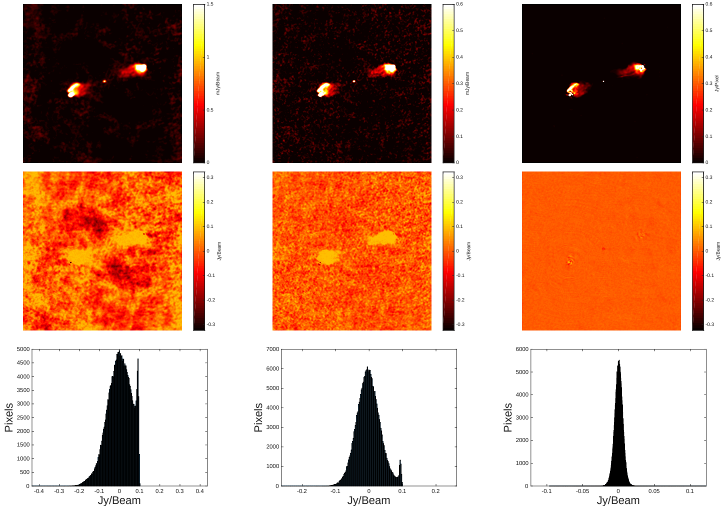


(a) CLEAN (natural) (b) CLEAN (uniform) (c) PURIFY

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

PURIFY reconstruction

VLA observation of Cygnus A



(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

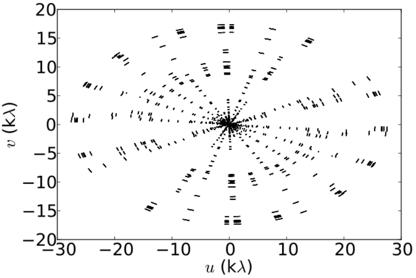
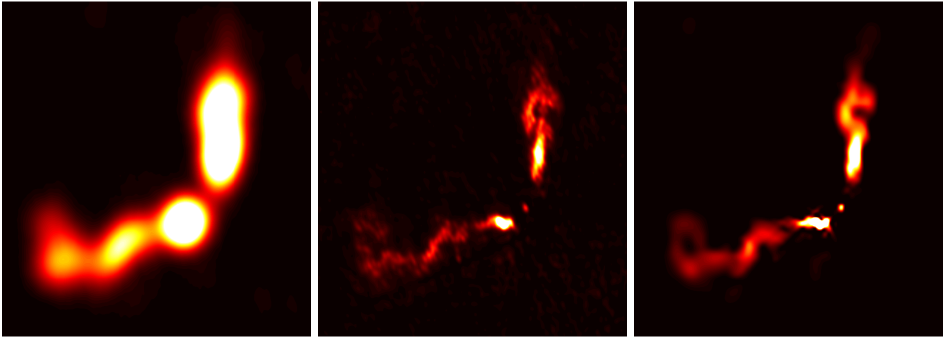


Figure: VLA visibility coverage for PKS J0334-39

PURIFY reconstruction

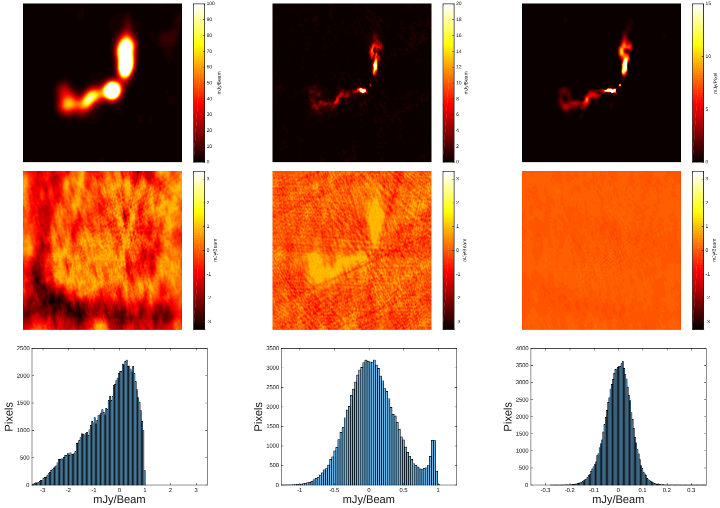
ATCA observation of PKS J0334-39



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

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

PURIFY reconstruction
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

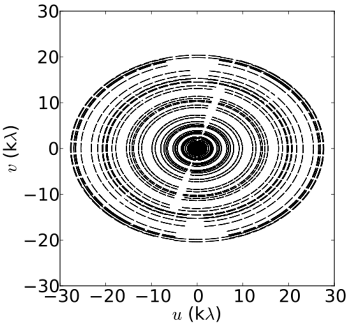
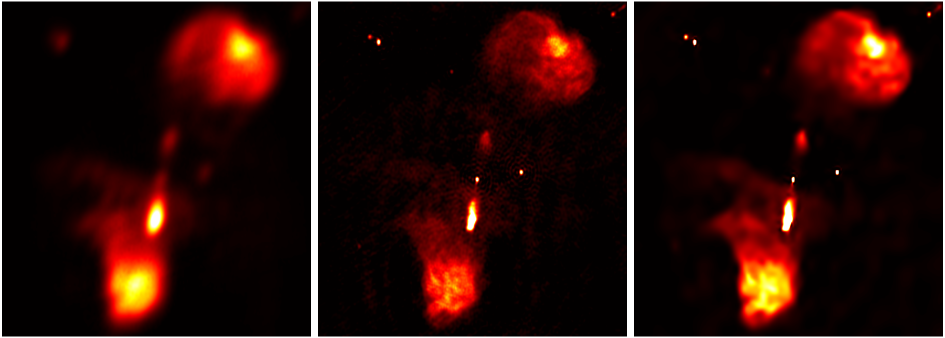


Figure: ATCA visibility coverage for Cygnus A

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

ATCA observation of PKS J0116-473

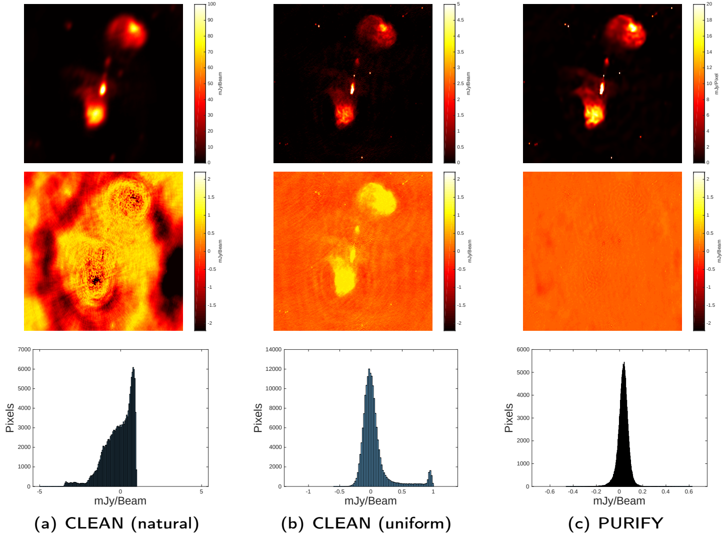


Figure: PKS J0116-473 recovered images and residuals (Pratley, McEwen, *et al.* 2016)

Distributed and parallelised convex optimisation

- Solve resulting convex optimisation problems by proximal splitting.
- Block inexact ADMM algorithm to split data and measurement operator: (Carrillo, McEwen & Wiaux 2014; Onose, Carrillo, Repetti, McEwen, *et al.* 2016)

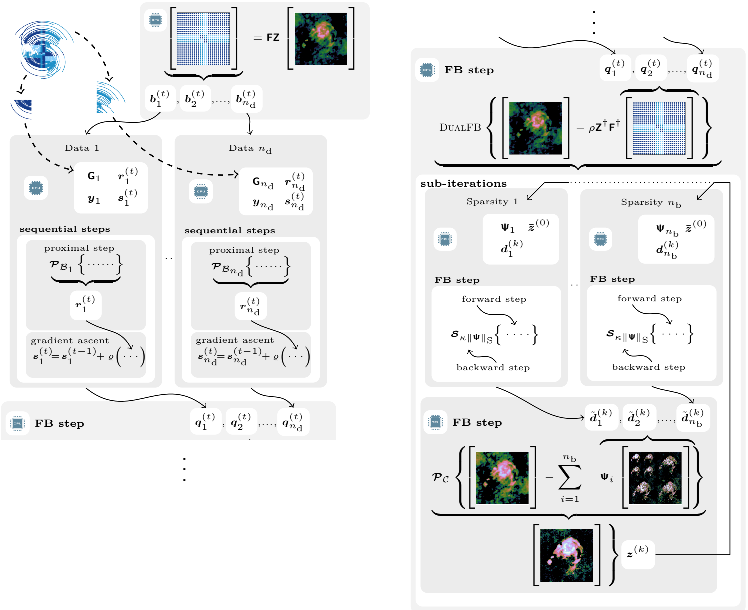
$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_{n_\mathrm{d}} \end{bmatrix}, \quad \boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_1 \\ \vdots \\ \boldsymbol{\Phi}_{n_\mathrm{d}} \end{bmatrix} = \begin{bmatrix} \mathsf{G}_1 \mathsf{M}_1 \\ \vdots \\ \mathsf{G}_{n_\mathrm{d}} \mathsf{M}_{n_\mathrm{d}} \end{bmatrix} \mathsf{FZ} .$$

Distributed and parallelised convex optimisation

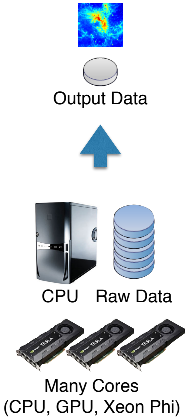
- Solve resulting convex optimisation problems by proximal splitting.
- Block inexact ADMM algorithm to split data and measurement operator:
  (Carrillo, McEwen & Wiaux 2014; Onose, Carrillo, Repetti, McEwen, *et al.* 2016)

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_{n_\mathrm{d}} \end{bmatrix}, \quad \boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_1 \\ \vdots \\ \boldsymbol{\Phi}_{n_\mathrm{d}} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1 \mathbf{M}_1 \\ \vdots \\ \mathbf{G}_{n_\mathrm{d}} \mathbf{M}_{n_\mathrm{d}} \end{bmatrix} \mathbf{F}\mathbf{Z} .$$

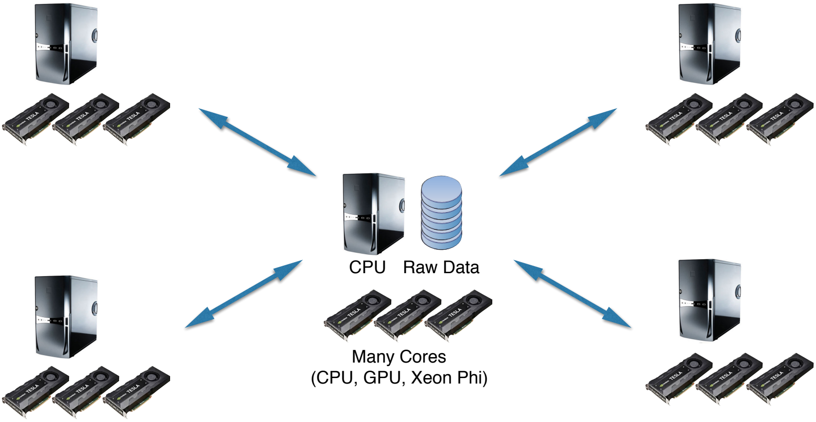
Distributed and parallelised convex optimisation



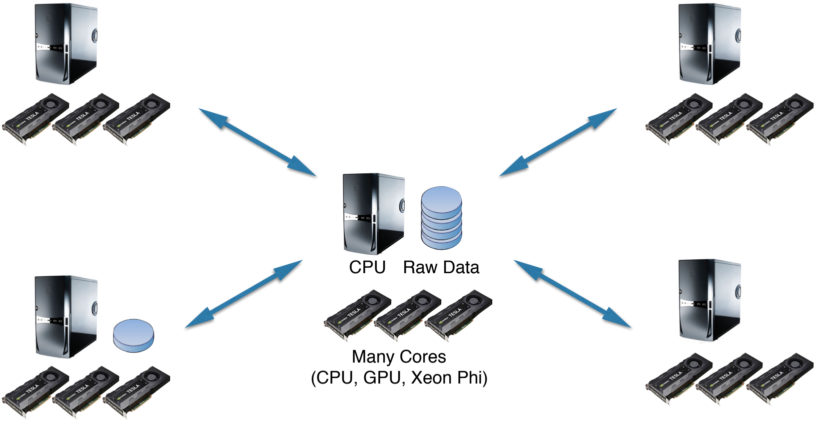
Standard algorithms



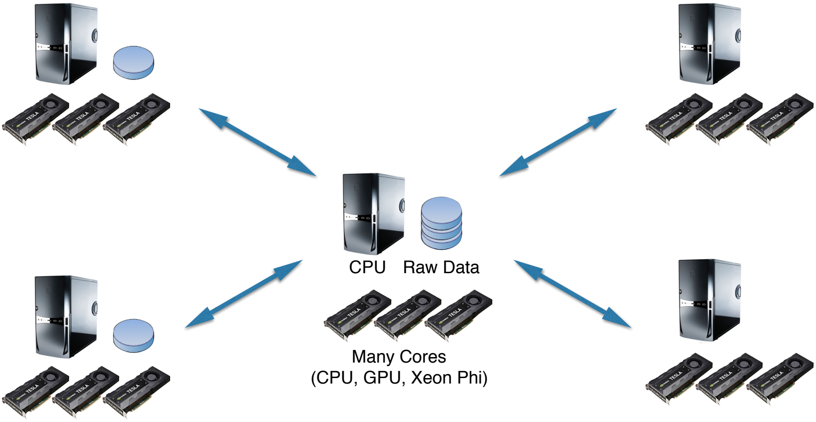
Highly distributed and parallelised algorithms



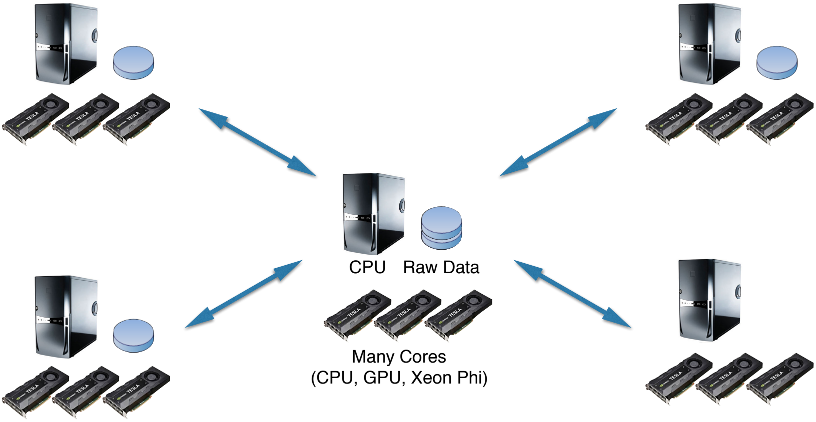
Highly distributed and parallelised algorithms



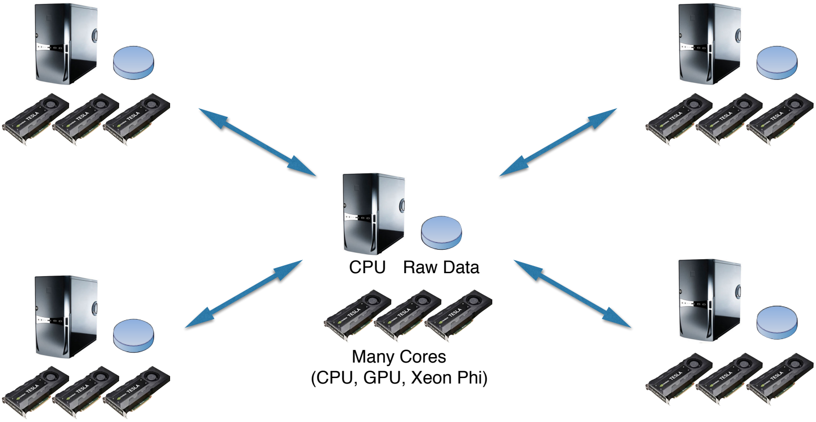
Highly distributed and parallelised algorithms



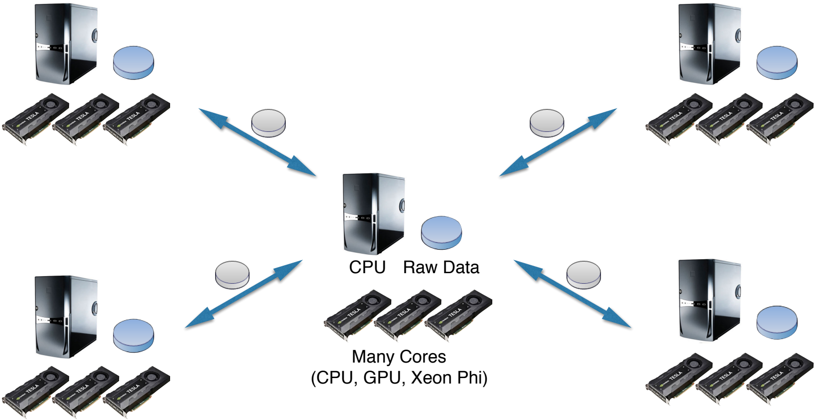
Highly distributed and parallelised algorithms



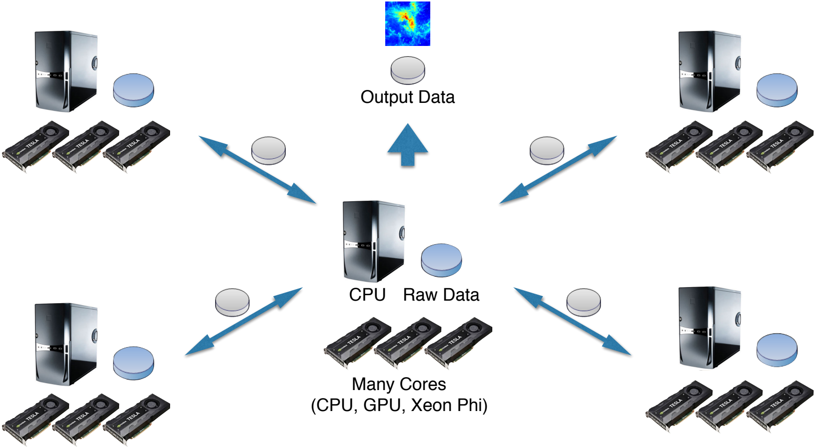
Highly distributed and parallelised algorithms



Highly distributed and parallelised algorithms



Highly distributed and parallelised algorithms



Outline

1. A unified framework for radio interferometric imaging
   - Bayesian inference
   - Regularisation
   - Compressive sensing
2. Compressive sensing for SKA imaging
   - PURIFY
   - Reconstruction fidelity
   - Scaling to big-data
3. Uncertainty quantification
   - Proximal MCMC
   - Compressive sensing with Bayesian credible intervals
   - Hypothesis testing

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_ℓ.
  - Sample from conditional distributions:

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

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

Sampling the full posterior distribution

Markov Chain Monte Carlo (MCMC)

- Alternative is to sample full posterior distribution $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y})$.

> $\Rightarrow$ 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_ℓ.
  - Sample from conditional distributions:

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

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

Sampling the full posterior distribution

Markov Chain Monte Carlo (MCMC)

- Alternative is to sample full posterior distribution $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y})$.

$\Rightarrow$ 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_ℓ.
  - Sample from conditional distributions:

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

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

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

- Alternative is to sample full posterior distribution $\mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y})$.

  $\Rightarrow$ 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_ℓ.
  - Sample from conditional distributions:

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

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

MCMC sampling with gradients

Langevin dynamics

- 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 $g(x)$ 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 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 sparse priors.

MCMC sampling with gradients

Langevin dynamics

- 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 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 sparse priors.

MCMC sampling with gradients

Langevin dynamics

- 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 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 sparse priors.

MCMC sampling with gradients

Langevin dynamics

- 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 π 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.

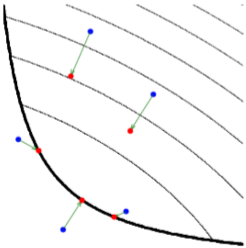


Figure: Illustration of proximity operator [Credit: Parikh & Boyd (2013)]

Proximal MCMC

Moreau approximation

- Follow Pereyra (2016a) and consider Moreau approximation of π:

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

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

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

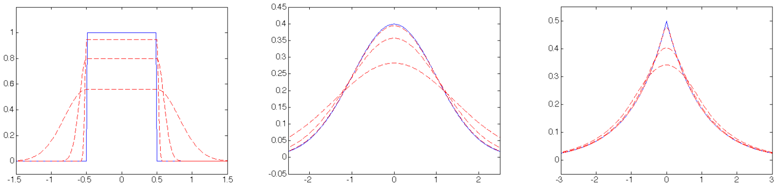


Figure: Illustration of Moreau approximations [Credit: Pereyra (2016a)]

Proximal MCMC
Moreau approximation

- Follow Pereyra (2016a) and consider Moreau approximation of π:

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

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

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

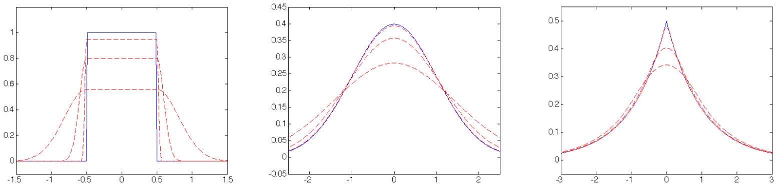


Figure: Illustration of Moreau approximations [Credit: Pereyra (2016a)]

Proximal MCMC

Proximal-MALA in the synthesis and analysis framework

Proximal Metropolis adjusted Langevin algorithm (P-MALA)

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

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

- Discretise and apply Moreau approximation:

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

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

- Metropolis-Hastings accept-reject step.

Need to compute $\mathrm{prox}_g^{\delta/2}$ for problem (Cai, Pereyra & McEwen, in prep.):

$$\simeq \mathrm{prox}_{\lambda\|\cdot\|_1}^{\delta/2}\Big(\boldsymbol{\alpha} - \delta\boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha} - \boldsymbol{y})\Big)$$

Synthesis framework

$$\simeq \mathrm{prox}_{\lambda\|\boldsymbol{\Psi}^\dagger\cdot\|_1}^{\delta/2}\Big(\boldsymbol{x} - \delta\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})\Big)$$

Analysis framework

Proximal MCMC

Proximal-MALA in the synthesis and analysis framework

Proximal Metropolis adjusted Langevin algorithm (P-MALA)

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

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

- Discretise and apply Moreau approximation:

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

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

- Metropolis-Hastings accept-reject step.

Need to compute $\mathrm{prox}_g^{\delta/2}$ for problem (Cai, Pereyra & McEwen, in prep.):

$$\simeq \mathrm{prox}^{\delta/2}_{\lambda\|\cdot\|_1}\Big(\boldsymbol{\alpha} - \delta\boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha} - \boldsymbol{y})\Big)$$

Synthesis framework

$$\simeq \mathrm{prox}^{\delta/2}_{\lambda\|\boldsymbol{\Psi}^\dagger\cdot\|_1}\Big(\boldsymbol{x} - \delta\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})\Big)$$

Analysis framework

Proximal MCMC

Proximal-MALA in the synthesis and analysis framework

Proximal Metropolis adjusted Langevin algorithm (P-MALA)

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

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

- Discretise and apply Moreau approximation:

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

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

- Metropolis-Hastings accept-reject step.

Need to compute $\mathrm{prox}_g^{\delta/2}$ for problem (Cai, Pereyra & McEwen, in prep.):

$$\simeq \mathrm{prox}_{\lambda\|\cdot\|_1}^{\delta/2}\Big(\boldsymbol{\alpha} - \delta\boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha} - \boldsymbol{y})\Big)$$

Synthesis framework

$$\simeq \mathrm{prox}_{\lambda\|\boldsymbol{\Psi}^\dagger\cdot\|_1}^{\delta/2}\Big(\boldsymbol{x} - \delta\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})\Big)$$

Analysis framework

Proximal MCMC

Proximal-MALA in the synthesis and analysis framework

Proximal Metropolis adjusted Langevin algorithm (P-MALA)

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

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

- Discretise and apply Moreau approximation:

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

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

- Metropolis-Hastings accept-reject step.

Need to compute $\mathrm{prox}_g^{\delta/2}$ for problem (Cai, Pereyra & McEwen, in prep.):

$$\simeq \mathrm{prox}_{\lambda\|\cdot\|_1}^{\delta/2}\Big(\boldsymbol{\alpha} - \delta\boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha} - \boldsymbol{y})\Big)$$

Synthesis framework

$$\simeq \mathrm{prox}_{\lambda\|\boldsymbol{\Psi}^\dagger\cdot\|_1}^{\delta/2}\Big(\boldsymbol{x} - \delta\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})\Big)$$

Analysis framework

Proximal MCMC

Proximal-MALA in the synthesis and analysis framework

Proximal Metropolis adjusted Langevin algorithm (P-MALA)

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

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

- Discretise and apply Moreau approximation:

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

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

- Metropolis-Hastings accept-reject step.

Need to compute $\mathrm{prox}_g^{\delta/2}$ for problem (Cai, Pereyra & McEwen, in prep.):

$$\simeq \mathrm{prox}_{\lambda\|\cdot\|_1}^{\delta/2}\Big(\boldsymbol{\alpha} - \delta\boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha} - \boldsymbol{y})\Big)$$

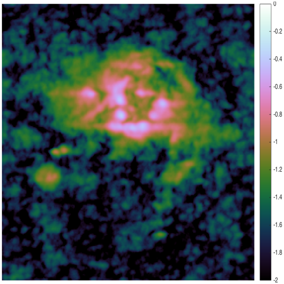
Synthesis framework

$$\simeq \mathrm{prox}_{\lambda\|\boldsymbol{\Psi}^\dagger\cdot\|_1}^{\delta/2}\Big(\boldsymbol{x} - \delta\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})\Big)$$

Analysis framework

Proximal MCMC

Preliminary results on simulations

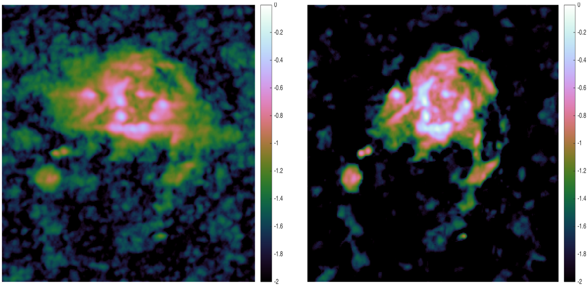


(a) Dirty image

Figure: HII region of M31

Proximal MCMC

Preliminary results on simulations

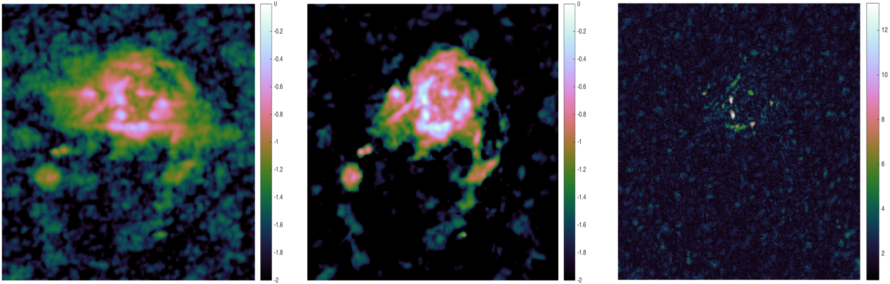


(a) Dirty image (b) Mean recovered image

Figure: HII region of M31

Proximal MCMC

Preliminary results on simulations

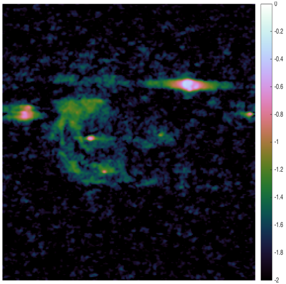


(a) Dirty image (b) Mean recovered image (c) Standard deviation image

Figure: HII region of M31

Proximal MCMC

Preliminary results on simulations

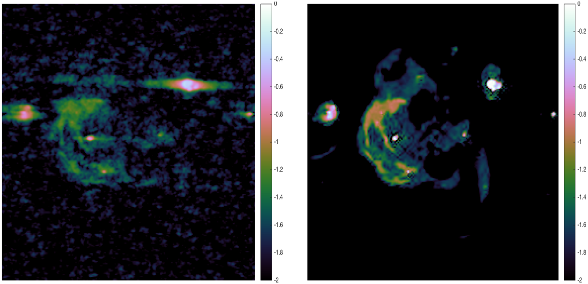


(a) Dirty image

Figure: Supernova remnant W28

Proximal MCMC

Preliminary results on simulations

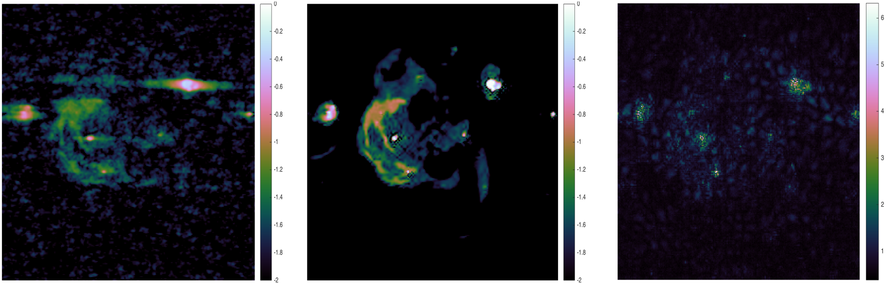


(a) Dirty image (b) Mean recovered image

Figure: Supernova remnant W28

Proximal MCMC

Preliminary results on simulations

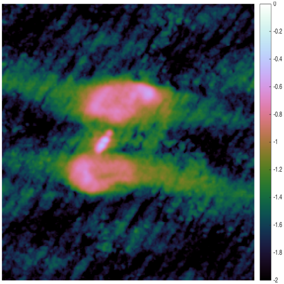


(a) Dirty image (b) Mean recovered image (c) Standard deviation image

Figure: Supernova remnant W28

Proximal MCMC

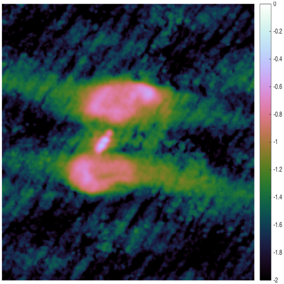
Preliminary results on simulations



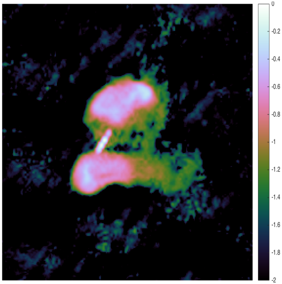
(a) Dirty image

Figure: 3C288

Proximal MCMC
Preliminary results on simulations



(a) Dirty image

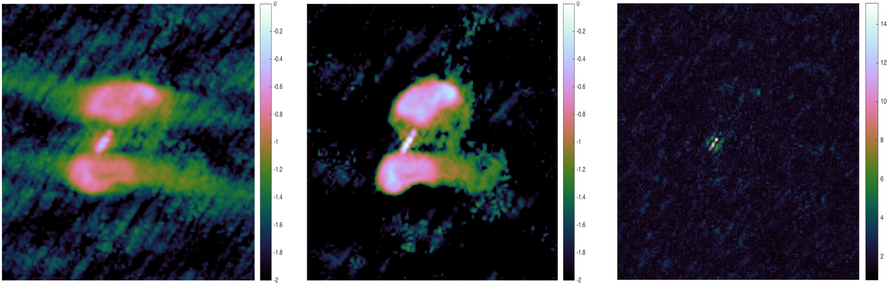


(b) Mean recovered image

Figure: 3C288

Proximal MCMC

Preliminary results on simulations



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

Figure: 3C288

Bayesian credible regions for compressive sensing

- Combine error estimation with fast sparse regularisation (*cf.* compressive sensing).
- Let C_α denote the highest posterior density (HPD) Bayesian credible region with confidence level $(1-\alpha)\%$ defined by posterior iso-contour: $C_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$.
- Analytic approximation $\tilde{\gamma}_\alpha = g(\boldsymbol{x}^\star) + N(\tau_\alpha + 1)$ (Pereyra 2016b).
- Compute $\boldsymbol{x}^\star$ by sparse regularisation and estimate local Bayesian credible intervals.

Bayesian credible regions for compressive sensing

- Combine error estimation with fast sparse regularisation (*cf.* compressive sensing).
- Let C_α denote the highest posterior density (HPD) Bayesian credible region with confidence level $(1-\alpha)\%$ defined by posterior iso-contour: $C_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$.
- Analytic approximation $\tilde{\gamma}_\alpha = g(\boldsymbol{x}^\star) + N(\tau_\alpha + 1)$ (Pereyra 2016b).
- Compute $x^\star$ by sparse regularisation and estimate local Bayesian credible intervals.

Bayesian credible regions for compressive sensing

- Combine error estimation with fast sparse regularisation (*cf.* compressive sensing).
- Let C_α denote the highest posterior density (HPD) Bayesian credible region with confidence level $(1-\alpha)\%$ defined by posterior iso-contour: $C_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$.
- Analytic approximation $\tilde{\gamma}_\alpha = g(\boldsymbol{x}^\star) + N(\tau_\alpha + 1)$ (Pereyra 2016b).
- Compute $\boldsymbol{x}^\star$ by sparse regularisation and estimate local Bayesian credible intervals.

Bayesian credible regions for compressive sensing

- Combine error estimation with fast sparse regularisation (*cf.* compressive sensing).
- Let C_α denote the highest posterior density (HPD) Bayesian credible region with confidence level $(1-\alpha)\%$ defined by posterior iso-contour: $C_\alpha = \{\boldsymbol{x} : g(\boldsymbol{x}) \leq \gamma_\alpha\}$.
- Analytic approximation $\tilde{\gamma}_\alpha = g(\boldsymbol{x}^\star) + N(\tau_\alpha + 1)$ (Pereyra 2016b).
- Compute $\boldsymbol{x}^\star$ by sparse regularisation and estimate local Bayesian credible intervals.

Local Bayesian credible intervals for sparse reconstruction (Cai, Pereyra & McEwen, in prep.)

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).

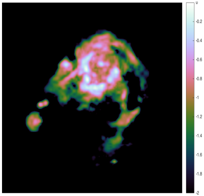
Given $\tilde{\gamma}_\alpha$ and $\boldsymbol{x}^\star$, compute the credible interval by

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

where

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

Bayesian credible regions
Preliminary results on simulations

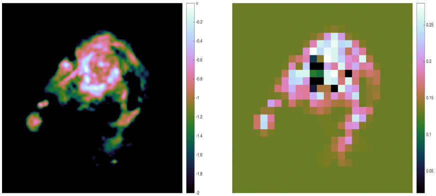


(a) Recovered image

Figure: HII region of M31

Bayesian credible regions

Preliminary results on simulations



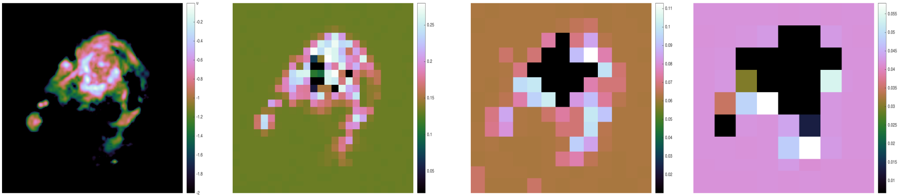
(a) Recovered image

(b) Credible intervals for regions of size 10×10

Figure: HII region of M31

Bayesian credible regions

Preliminary results on simulations



(a) Recovered image

(b) Credible intervals for regions of size 10×10

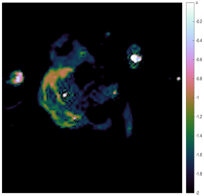
(c) Credible intervals for regions of size 20×20

(d) Credible intervals for regions of size 30×30

Figure: HII region of M31

Bayesian credible regions

Preliminary results on simulations

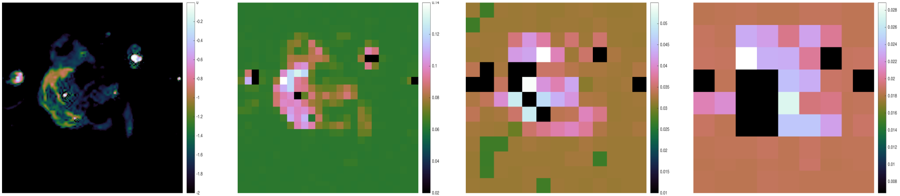


(a) Recovered image

Figure: Supernova remnant W28

Bayesian credible regions

Preliminary results on simulations

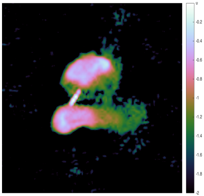


(a) Recovered image (b) Credible intervals for regions of size 10×10 (c) Credible intervals for regions of size 20×20 (d) Credible intervals for regions of size 30×30

Figure: Supernova remnant W28

Bayesian credible regions

Preliminary results on simulations

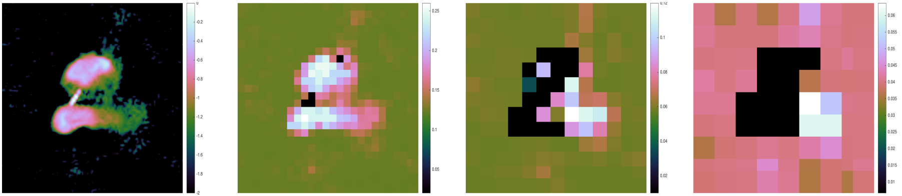


(a) Recovered image

Figure: 3C288

Bayesian credible regions

Preliminary results on simulations



(a) Recovered image

(b) Credible intervals for regions of size 10×10

(c) Credible intervals for regions of size 20×20

(d) Credible intervals for regions of size 30×30

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 $\boldsymbol{x}_\star$.
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 artefact 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

- 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 $\boldsymbol{x}_\star$.
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 artefact 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

- 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 $\boldsymbol{x}_\star$.
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 artefact 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

- 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 $\boldsymbol{x}_\star$.
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 artefact 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

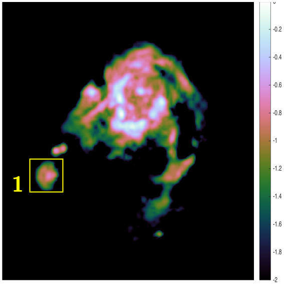
- Is structure in an image physical or an artefact?
- Can we make precise statistical statements?
- Perform hypothesis tests using Bayesian credible regions (Pereyra 2016b).

Hypothesis testing of physical structure

1. Cut out region containing structure of interest from recovered image $\boldsymbol{x}_\star$.
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 artefact 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

Preliminary results on simulations

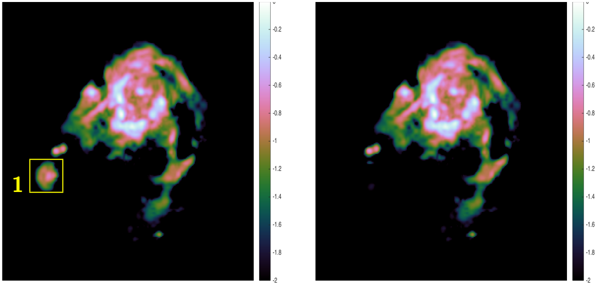


(a) Recovered image

Figure: HII region of M31

Hypothesis testing

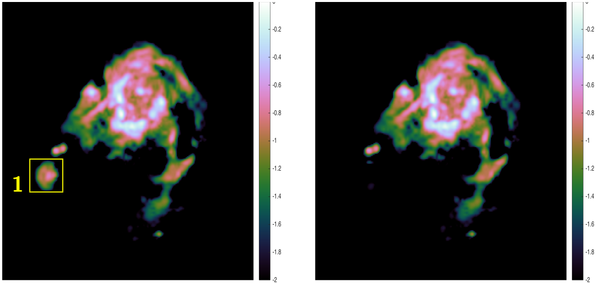
Preliminary results on simulations



(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

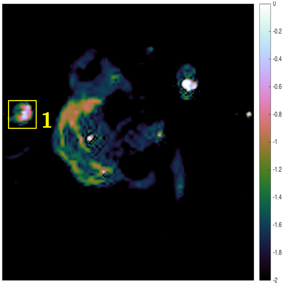
Figure: HII region of M31

Reject null hypothesis

⇒ structure physical

Hypothesis testing

Preliminary results on simulations

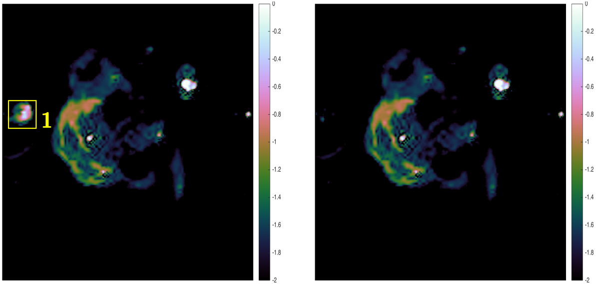


(a) Recovered image

Figure: Supernova remnant W28

Hypothesis testing

Preliminary results on simulations



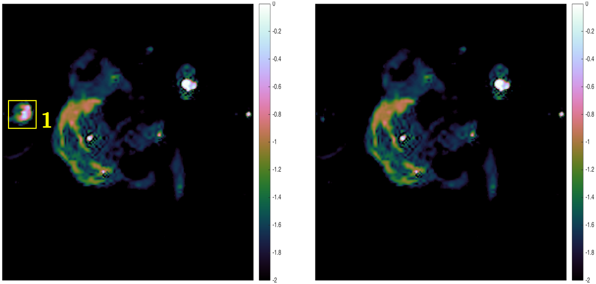
(a) Recovered image

(b) Surrogate with region removed

Figure: Supernova remnant W28

Hypothesis testing

Preliminary results on simulations



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

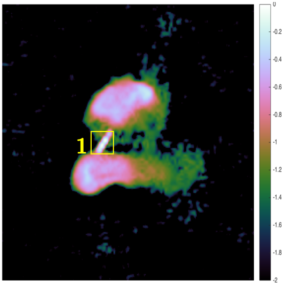
Figure: Supernova remnant W28

Reject null hypothesis

⇒ structure physical

Hypothesis testing

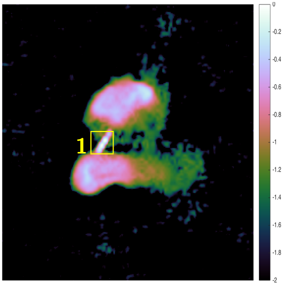
Preliminary results on simulations



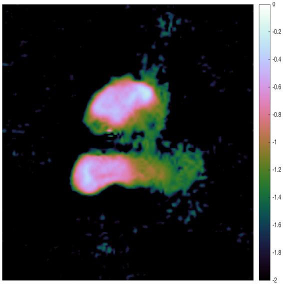
(a) Recovered image

Figure: 3C288

Hypothesis testing
Preliminary results on simulations



(a) Recovered image

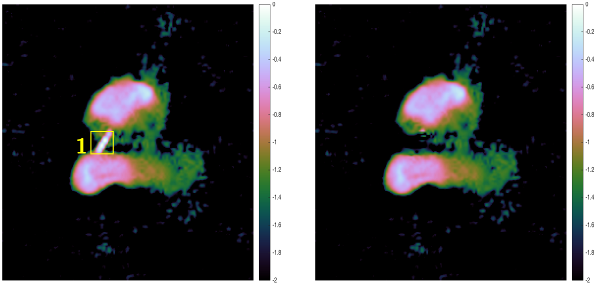


(b) Surrogate with region removed

Figure: 3C288

Hypothesis testing

Preliminary results on simulations



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

Figure: 3C288

Reject null hypothesis

⇒ structure physical

Conclusions

1. Unified framework for interferometric imaging.

   Sparse priors (*cf.* compressive sensing) shown to be highly effective and scalable to big-data.

   PURIFY package provides robust framework for imaging interferometric observations (`http://basp-group.github.io/purify/`).

2. Seek statistical interpretation to recover error information.

   Proximal MCMC sampling can support sparse priors in full statistical framework.

   Combine error estimation with fast sparse regularisation (*cf.* compressive sensing):

   - Recover Bayesian credible regions.
   - Perform hypothesis testing to test whether structure physical.

Supported by:

EPSRC Engineering and Physical Sciences Research Council

Science & Technology Facilities Council

Conclusions

1. Unified framework for interferometric imaging.

   Sparse priors (*cf.* compressive sensing) shown to be highly effective and scalable to big-data.

   PURIFY package provides robust framework for imaging interferometric observations (http://basp-group.github.io/purify/).

2. Seek statistical interpretation to recover error information.

   Proximal MCMC sampling can support sparse priors in full statistical framework.

   Combine error estimation with fast sparse regularisation (*cf.* compressive sensing):

   - Recover Bayesian credible regions.
   - Perform hypothesis testing to test whether structure physical.

Supported by:

EPSRC Engineering and Physical Sciences Research Council

Science & Technology Facilities Council

Extra Slides

Compressive sensing | Analysis vs synthesis | Bayesian interpretations

PURIFY reconstructions

Extra Slides

Compressive sensing

An introduction to compressive sensing

Operator description

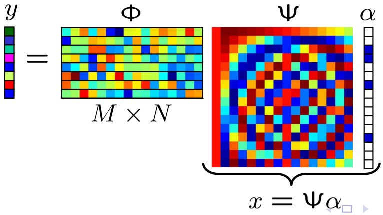
- Linear operator (linear algebra) representation of signal decomposition:

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

- Linear operator (linear algebra) representation of measurement:

$$y_i = \langle x, \Phi_j \rangle \quad \rightarrow \quad \boldsymbol{y} = \begin{pmatrix} - & \Phi_0 & - \\ - & \Phi_1 & - \\ & \vdots & \end{pmatrix} \boldsymbol{x} \quad \rightarrow \quad \boxed{\boldsymbol{y} = \boldsymbol{\Phi} \boldsymbol{x}}$$

- Putting it together: $\boxed{\boldsymbol{y} = \boldsymbol{\Phi} \boldsymbol{x} = \boldsymbol{\Phi} \boldsymbol{\Psi} \boldsymbol{\alpha}}$



An introduction to compressive sensing

Promoting sparsity via ℓ_1 minimisation

- Ill-posed inverse problem:

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

- Solve by imposing a regularising prior that the signal to be recovered is sparse in Ψ, *i.e.* solve the following ℓ_0 optimisation problem:

$$\boxed{\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \text{ subject to } \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon},$$

where the signal is synthesised by $\boldsymbol{x}^\star = \Psi\boldsymbol{\alpha}^\star$.

- Recall norms given by:

$$\|\boldsymbol{\alpha}\|_0 = \text{no. non-zero elements} \qquad \|\boldsymbol{\alpha}\|_1 = \sum_i |\boldsymbol{\alpha}_i| \qquad \|\boldsymbol{\alpha}\|_2^2 = \sum_i |\boldsymbol{\alpha}_i|^2$$

- Solving this problem is difficult (combinatorial).

- Instead, solve the ℓ_1 optimisation problem (convex):

$$\boxed{\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ subject to } \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon}.$$

An introduction to compressive sensing

Union of subspaces

- Space of sparse vectors given by the union of subspaces aligned with the coordinate axes.

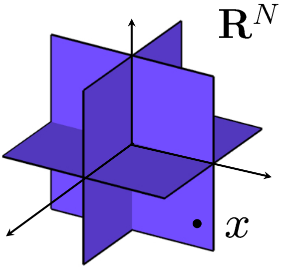


Figure: Space of the sparse vectors **[Credit: Baraniuk]**

An introduction to compressive sensing
Restricted isometry property (RIP)

- Solutions of ℓ_0 and ℓ_1 problems often the same.
- Restricted isometry property (RIP):
$$(1-\delta_{2K})\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2 \leq \|\boldsymbol{\Theta}\boldsymbol{x}_1-\boldsymbol{\Theta}\boldsymbol{x}_2\|_2^2 \leq (1+\delta_{2K})\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2^2\ ,$$
for K-sparse $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, where $\boldsymbol{\Theta} = \Phi\Psi$.
- Measurement must preserve geometry of sets of sparse vectors.

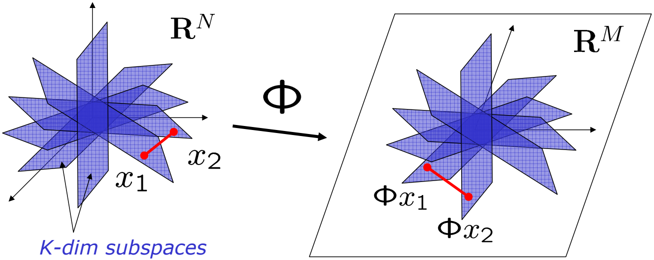


Figure: Measurement must preserve geometry of sets of sparse vectors. **[Credit: Baraniuk]**

An introduction to compressive sensing

Intuition

- Solutions of ℓ_0 and ℓ_1 problems often the same.
- Geometry of ℓ_0, ℓ_2 and ℓ_1 problems.

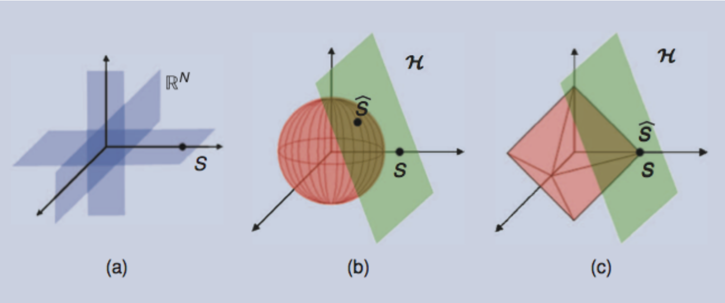


Figure: Geometry of (a) ℓ_0 (b) ℓ_2 and (c) ℓ_1 problems. **[Credit: Baraniuk (2007)]**

An introduction to compressive sensing

Sparsity and coherence

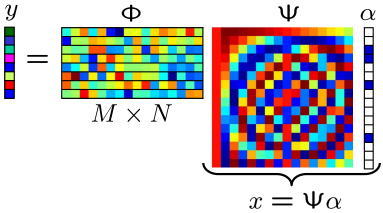
- In the absence of noise, compressed sensing is exact!
- Number of measurements required to achieve exact reconstruction is given by

$$\boxed{M \geq c\mu^2 K \log N},$$

where K is the sparsity and N the dimensionality.

- The coherence between the measurement and sparsity basis is given by

$$\boxed{\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle|}.$$



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):

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

analysis

- Contrast with synthesis-based approach:

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

synthesis

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

Analysis vs synthesis

Comparison

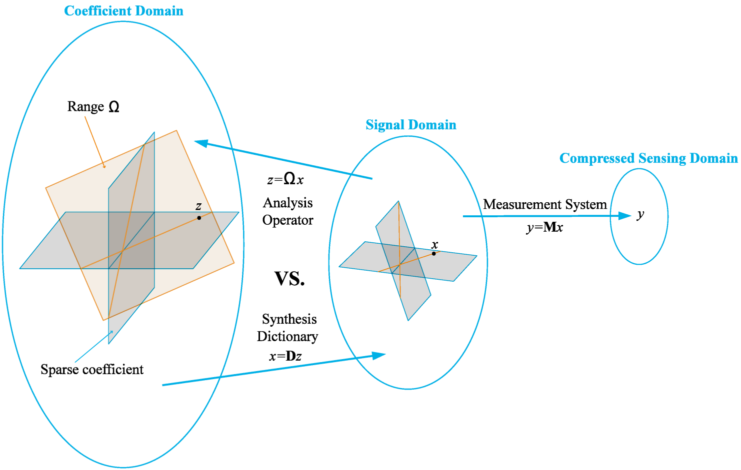


Figure: Analysis- and synthesis-based approaches **[Credit: Nam *et al.* (2012)]**.

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

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

synthesis

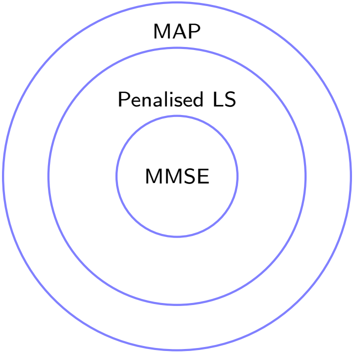
- One possible Bayesian interpretation!

- 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

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

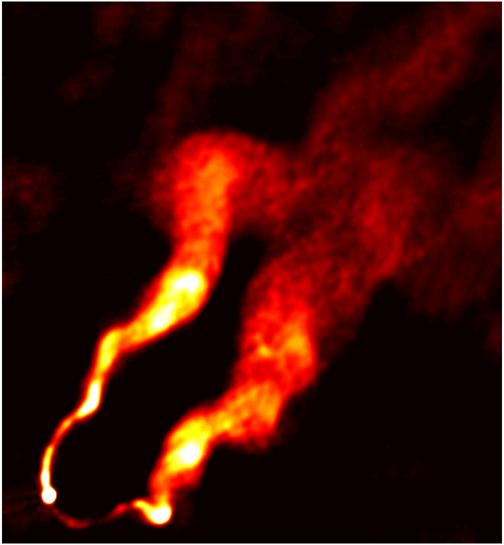
analysis

- 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

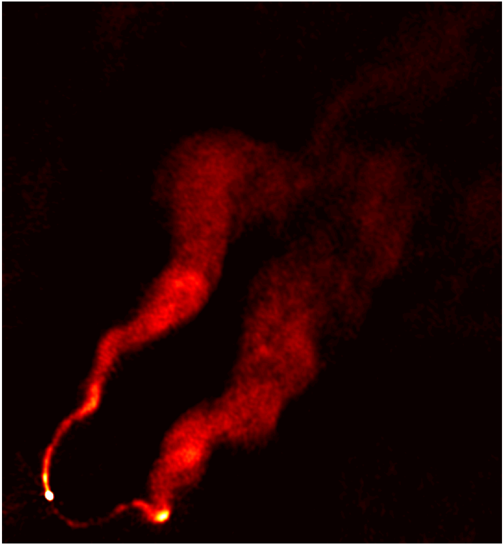
PURIFY reconstructions

CLEAN (natural) reconstruction
VLA observation of 3C129

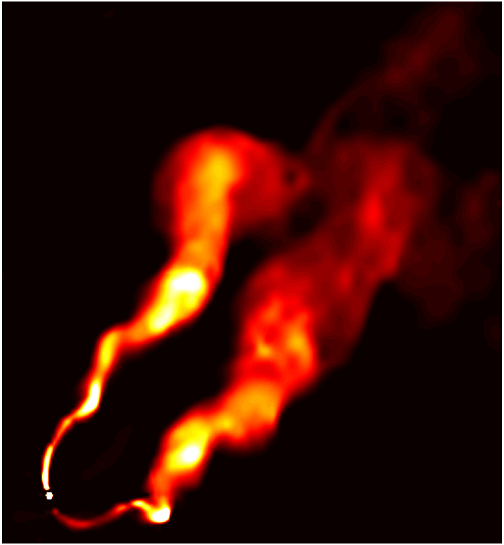


CLEAN (uniform) reconstruction

VLA observation of 3C129

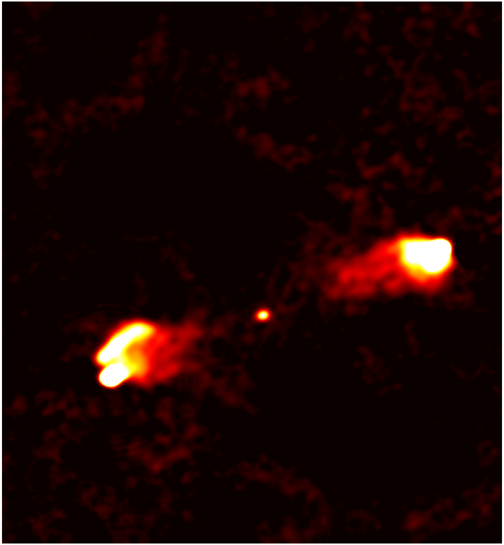


PURIFY reconstruction
VLA observation of 3C129



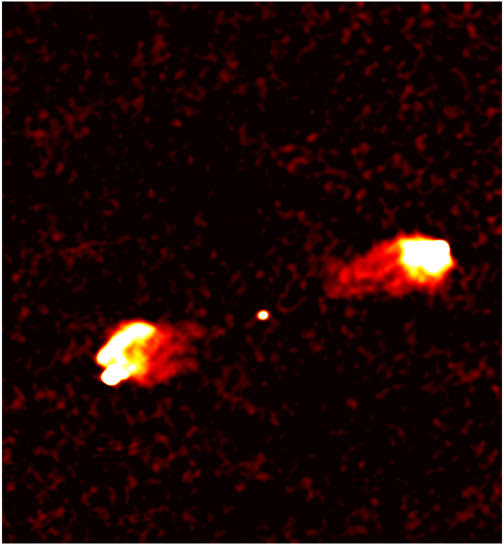
CLEAN (natural) reconstruction

VLA observation of Cygnus A



CLEAN (uniform) reconstruction

VLA observation of Cygnus A

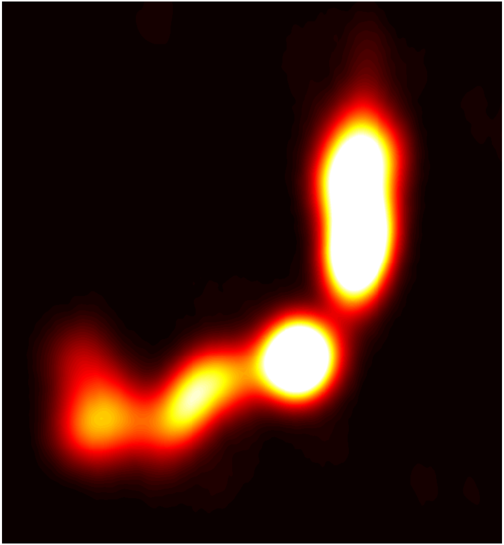


PURIFY reconstruction

VLA observation of Cygnus A

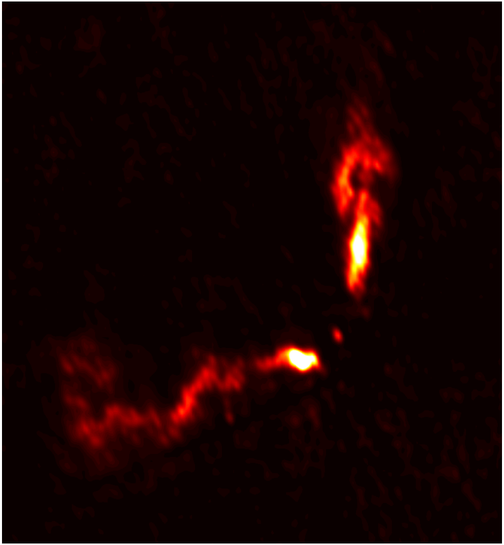
CLEAN (natural) reconstruction

ATCA observation of PKS J0334-39



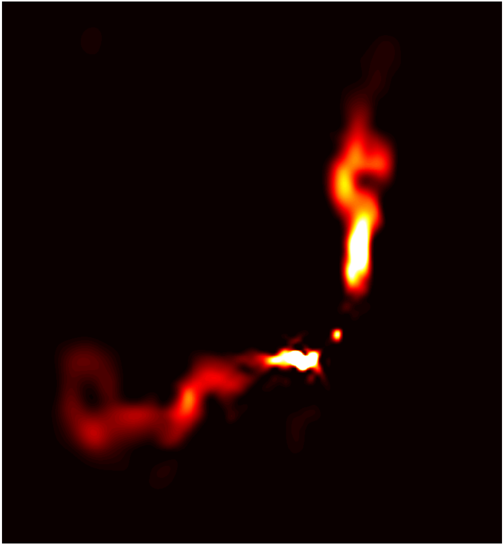
CLEAN (uniform) reconstruction

ATCA observation of PKS J0334-39



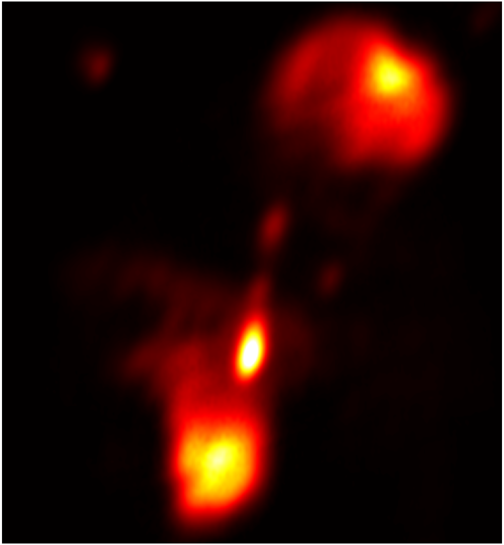
PURIFY reconstruction

ATCA observation of PKS J0334-39



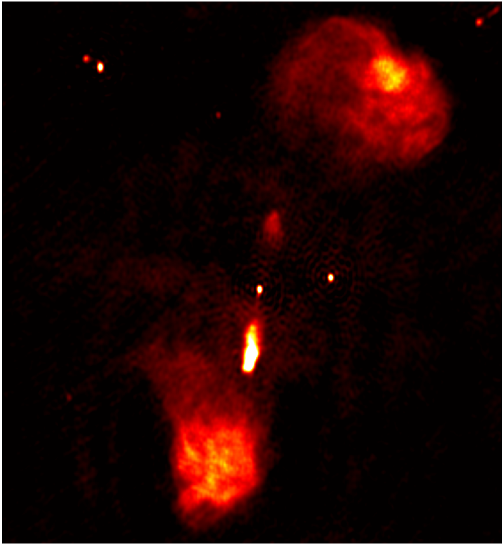
CLEAN (natural) reconstruction

ATCA observation of PKS J0116-473



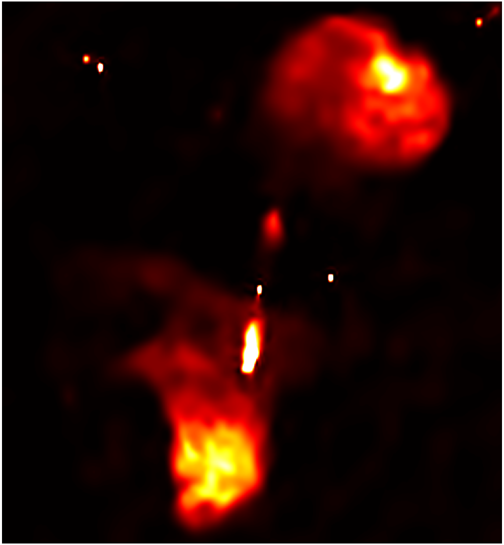
CLEAN (uniform) reconstruction

ATCA observation of PKS J0116-473



PURIFY reconstruction

ATCA observation of PKS J0116-473



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