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



Bayesian interpretations of interferometric imaging techniques





Sparse regularisation with Bayesian credible regions

▲□▶ ▲□▶ ★ □▶ ★ □▶ - □ - つくぐ

Outline



Bayesian interpretations of interferometric imaging techniques

Proximal MCMC sampling

3 Sparse regularisation with Bayesian credible regions

▲□▶ ▲□▶ ★ □▶ ★ □▶ - □ - つくぐ

Radio interferometric inverse problem

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

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

where y are the measured visibilities, Φ is the linear measurement operator, x is the underlying image and n is instrumental noise.

- - primary beam A of the telescope;
 - Fourier transform F;
 - convolutional de-gridding G to interpolate to continuous uv-coordinates;
 - direction-dependent effects (DDEs)...

・ロト ・四ト ・ヨト ・ヨト

Radio interferometric inverse problem

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

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

where y are the measured visibilities, Φ is the linear measurement operator, x is the underlying image and n is instrumental noise.

- Measurement operator, *e.g.* $\Phi = GFA$, 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)...

イロト イポト イヨト イヨト 二日

Radio interferometric inverse problem

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

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

where y are the measured visibilities, Φ is the linear measurement operator, x is the underlying image and n is instrumental noise.

• Measurement operator, *e.g.* $\Phi = GFA$, 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.

イロト イポト イヨト イヨト 二日

Bayesian evolution



・ロト ・四ト ・ヨト ・ヨト

э

Bayesian inference

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

Bayes to the rescue:

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





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

• 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 y (visibilities) and model M (interferometric telescope with Gaussian noise), we want a full probabilistic description of our knowledge of the underlying sky image x.
- Bayes to the rescue:

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





• Bayes theorem in words:

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

- 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 y (visibilities) and model M (interferometric telescope with Gaussian noise), we want a full probabilistic description of our knowledge of the underlying sky image x.
- Bayes to the rescue:

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





• Bayes theorem in words:

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

- How do we perform Bayesian inference in practice?
 - \Rightarrow maximum a-posteriori (MAP) estimates and sampling approaches (MCMC)

(and many others)



Figure: Probability distribution to explore in 2D

イロト イヨト イヨト イヨト



Figure: Maximum a-posteriori (MAP) estimate

ヘロト ヘアト ヘビト ヘビト



Figure: Markov Chain Monte Carlo (MCMC) sampling

イロン イロン イヨン イヨン



Figure: Markov Chain Monte Carlo (MCMC) sampling

<ロ> (四) (四) (三) (三) (三)



Figure: Markov Chain Monte Carlo (MCMC) sampling

<ロ> (四) (四) (三) (三) (三)

- 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...

э

• Start with Bayes Theorem (ignore normalising evidence):

```
P(\boldsymbol{x} \,|\, \boldsymbol{y}) \propto P(\boldsymbol{y} \,|\, \boldsymbol{x}) P(\boldsymbol{x}) \,, \quad \textit{i.e. posterior} \propto \mathsf{likelihood} 	imes \mathsf{prior}
```

MAP estimator:

◆□▶ ◆□▶ ◆目▶ ◆目▶ 目 のへで

• Start with Bayes Theorem (ignore normalising evidence):

```
\mathrm{P}({m{x}}\,|\,{m{y}}) \propto \mathrm{P}({m{y}}\,|\,{m{x}})\mathrm{P}({m{x}})\,, \quad \textit{i.e.} \text{ posterior } \propto \mathsf{likelihood} 	imes \mathsf{prior}
```

Define likelihood (assuming Gaussian noise) and prior:

MAP estimator:

◆□▶ ◆□▶ ◆目▶ ◆目▶ 目 のへで

• Start with Bayes Theorem (ignore normalising evidence):

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

Define likelihood (assuming Gaussian noise) and prior:

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

Likelihood

MAP estimator:

イロト イロト イヨト イヨト 二日

• Start with Bayes Theorem (ignore normalising evidence):

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

Define likelihood (assuming Gaussian noise) and prior:

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

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

Prior

э

Likelihood

• MAP estimator:

• Start with Bayes Theorem (ignore normalising evidence):

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

Define likelihood (assuming Gaussian noise) and prior:

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

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

Likelihood

Prior

Consider log-posterior:

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

MAP estimator:

• Start with Bayes Theorem (ignore normalising evidence):

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

Define likelihood (assuming Gaussian noise) and prior:

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

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

イロト イポト イヨト イヨト 二日

Likelihood

Prior

Consider log-posterior:

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

• MAP estimator:

$$oldsymbol{x}_{ ext{map}} = rg\max_{oldsymbol{x}} \Big[\log \mathrm{P}(oldsymbol{y} \,|\, oldsymbol{x}) \Big]$$

• Start with Bayes Theorem (ignore normalising evidence):

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

Define likelihood (assuming Gaussian noise) and prior:

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

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

Likelihood

Prior

Consider log-posterior:

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

• MAP estimator:

$$\boldsymbol{x}_{\text{map}} = \underset{\boldsymbol{x}}{\arg \max} \left[\log P(\boldsymbol{y} | \boldsymbol{x}) \right] = \underset{\boldsymbol{x}}{\arg \min} \left[\underbrace{ \left\| \boldsymbol{y} - \boldsymbol{\Phi} \boldsymbol{x} \right\|_{2}^{2}}_{\text{Data fidelity}} + \underbrace{ \lambda R(\boldsymbol{x})}_{\text{Regulariser}} \right]$$

CLEAN

Consider the sparse prior: $P(\boldsymbol{x}) \propto \exp\left(-\beta \|\boldsymbol{x}\|_{0}\right)$.

Corresponding MAP estimator is:

$$oldsymbol{x}_{ ext{clean}} = rgmin_{oldsymbol{x}} \Big[egin{smallmet} oldsymbol{y} - oldsymbol{\Phi} oldsymbol{x} igg|_2^2 + \lambda egin{smallmet} oldsymbol{x} igg|_0 \Big] .$$

(Laplace prior $\mathrm{P}(m{x}) \propto \exp\!\left(-eta \left\|m{x}
ight\|_1
ight)$ is good proxy; Wiaux *et al.* 2009)

MEM

Consider the entropic prior: $\mathrm{P}(m{x}) \propto \expig(-eta \, m{x}^\dagger \log m{x}ig).$

Corresponding MAP estimator is:

$$oldsymbol{x}_{ ext{mem}} = rgmin_{oldsymbol{x}} \left[egin{smallmed} oldsymbol{y} - oldsymbol{\Phi} oldsymbol{x} igg|_2^2 + \lambda \, oldsymbol{x}^\dagger \log oldsymbol{x}
ight].$$

CLEAN

Consider the sparse prior: $P(\boldsymbol{x}) \propto \exp\left(-\beta \|\boldsymbol{x}\|_{0}\right)$.

Corresponding MAP estimator is:

$$oldsymbol{x}_{ ext{clean}} = rgmin_{oldsymbol{x}} \Big[egin{array}{clean} oldsymbol{y} - oldsymbol{\Phi} oldsymbol{x} ig\|_{2}^{2} + \lambda egin{array}{clean} oldsymbol{x} ig\|_{0} \end{bmatrix}$$

(Laplace prior $P(\boldsymbol{x}) \propto \exp\left(-\beta \|\boldsymbol{x}\|_{1}\right)$ is good proxy; Wiaux *et al.* 2009)

MEM

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

Corresponding MAP estimator is:

$$oldsymbol{x}_{ ext{mem}} = rgmin_{oldsymbol{x}} \Big[egin{smallmet} oldsymbol{y} - oldsymbol{\Phi} oldsymbol{x} \Big\|_2^2 + \lambda \, oldsymbol{x}^\dagger \log oldsymbol{x} \Big]$$

CLEAN

Consider the sparse prior: $P(\boldsymbol{x}) \propto \exp\left(-\beta \|\boldsymbol{x}\|_{0}\right)$.

Corresponding MAP estimator is:

$$oldsymbol{x}_{ ext{clean}} = rgmin_{oldsymbol{x}} \Big[egin{smallmet} oldsymbol{y} - oldsymbol{\Phi} oldsymbol{x} ig\|_2^2 + \lambda egin{smallmet} oldsymbol{x} ig\|_0 \Big] \end{array}$$

(Laplace prior $P(\boldsymbol{x}) \propto \exp\left(-\beta \|\boldsymbol{x}\|_{1}\right)$ is good proxy; Wiaux *et al.* 2009)

MEM

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

Corresponding MAP estimator is:

$$oldsymbol{x}_{ ext{mem}} = rgmin_{oldsymbol{x}} \Big[oldsymbol{\|y - \Phi x\|}_2^2 + \lambda \, oldsymbol{x}^\dagger \log oldsymbol{x} \Big]$$

CLEAN

Consider the sparse prior: $P(\boldsymbol{x}) \propto \exp\left(-\beta \|\boldsymbol{x}\|_{0}\right)$.

Corresponding MAP estimator is:

$$oldsymbol{x}_{ ext{clean}} = rgmin_{oldsymbol{x}} \Big[egin{smallmet} oldsymbol{y} - oldsymbol{\Phi} oldsymbol{x} ig\|_2^2 + \lambda egin{smallmet} oldsymbol{x} ig\|_0 \Big] \end{array}$$

(Laplace prior $P(\boldsymbol{x}) \propto \exp\left(-\beta \|\boldsymbol{x}\|_{1}\right)$ is good proxy; Wiaux *et al.* 2009)

MEM

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

Corresponding MAP estimator is:

$$oldsymbol{x}_{ ext{mem}} = rgmin_{oldsymbol{x}} \Big[ig\| oldsymbol{y} - oldsymbol{\Phi} oldsymbol{x} ig\|_2^2 + \lambda \, oldsymbol{x}^\dagger \log oldsymbol{x} \Big]$$

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

CLEAN

Consider the sparse prior: $P(\boldsymbol{x}) \propto \exp\left(-\beta \|\boldsymbol{x}\|_{0}\right)$.

Corresponding MAP estimator is:

$$oldsymbol{x}_{ ext{clean}} = rgmin_{oldsymbol{x}} \Big[egin{smallmet} oldsymbol{y} - oldsymbol{\Phi} oldsymbol{x} ig\|_2^2 + \lambda egin{smallmet} oldsymbol{x} ig\|_0 \Big] \end{array}$$

(Laplace prior $P(\boldsymbol{x}) \propto \exp\left(-\beta \|\boldsymbol{x}\|_{1}\right)$ is good proxy; Wiaux *et al.* 2009)

MEM

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

Corresponding MAP estimator is:

$$oldsymbol{x}_{ ext{mem}} = rgmin_{oldsymbol{x}} \Big[oldsymbol{\|y - \Phi x\|}_2^2 + \lambda \, oldsymbol{x}^\dagger \log oldsymbol{x} \Big]$$

• Consider sparsifying representation (e.g. wavelet basis):

$$\boldsymbol{x} = \sum_{i} \boldsymbol{\Psi}_{i} \alpha_{i} = \begin{pmatrix} | \\ \boldsymbol{\Psi}_{0} \\ | \end{pmatrix} \alpha_{0} + \begin{pmatrix} | \\ \boldsymbol{\Psi}_{1} \\ | \end{pmatrix} \alpha_{1} + \cdots \quad \Rightarrow \quad \boldsymbol{x} = \boldsymbol{\Psi} \boldsymbol{\alpha}$$

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

$$\boldsymbol{x}_{\text{synthesis}} = \boldsymbol{\Psi} \times \operatorname*{arg\,min}_{\boldsymbol{\alpha}} \Big[\left\| \boldsymbol{y} - \boldsymbol{\Phi} \boldsymbol{\Psi} \boldsymbol{\alpha} \right\|_{2}^{2} + \lambda \left\| \boldsymbol{\alpha} \right\|_{1} \Big]$$

Synthesis framework

・ロト ・四ト ・ヨト ・ヨト

• Consider sparsifying representation (e.g. wavelet basis):

$$oldsymbol{x} = \sum_{i} \Psi_{i} lpha_{i} = \begin{pmatrix} | \\ \Psi_{0} \\ | \end{pmatrix} lpha_{0} + \begin{pmatrix} | \\ \Psi_{1} \\ | \end{pmatrix} lpha_{1} + \cdots \quad \Rightarrow \quad \boxed{oldsymbol{x} = \Psi lpha}$$

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

$$\boldsymbol{x}_{\text{synthesis}} = \boldsymbol{\Psi} \times \operatorname*{arg\,min}_{\boldsymbol{\alpha}} \Big[\left\| \boldsymbol{y} - \boldsymbol{\Phi} \boldsymbol{\Psi} \boldsymbol{\alpha} \right\|_{2}^{2} + \lambda \left\| \boldsymbol{\alpha} \right\|_{1} \Big]$$

Synthesis framework

・ロト ・四ト ・ヨト ・ヨト

• Consider sparsifying representation (e.g. wavelet basis):

$$\boldsymbol{x} = \sum_{i} \Psi_{i} \alpha_{i} = \begin{pmatrix} | \\ \Psi_{0} \\ | \end{pmatrix} \alpha_{0} + \begin{pmatrix} | \\ \Psi_{1} \\ | \end{pmatrix} \alpha_{1} + \cdots \quad \Rightarrow \quad \boxed{\boldsymbol{x} = \Psi \alpha}$$

- Recover (wavelet) coefficients α of image x.
- Consider the Laplacian prior on coefficients: $P(\alpha) \propto \exp(-\beta \|\alpha\|_1)$.

• Sparse synthesis regularisation problem:

$$\boldsymbol{x}_{\text{synthesis}} = \boldsymbol{\Psi} \times \argmin_{\boldsymbol{\alpha}} \left[\left\| \boldsymbol{y} - \boldsymbol{\Phi} \boldsymbol{\Psi} \boldsymbol{\alpha} \right\|_{2}^{2} + \lambda \left\| \boldsymbol{\alpha} \right\|_{1} \right]$$

Synthesis framework

・ロト ・四ト ・ヨト ・ヨト

э

• Consider sparsifying representation (e.g. wavelet basis):

$$oldsymbol{x} = \sum_{i} \Psi_{i} lpha_{i} = \begin{pmatrix} | \\ \Psi_{0} \\ | \end{pmatrix} lpha_{0} + \begin{pmatrix} | \\ \Psi_{1} \\ | \end{pmatrix} lpha_{1} + \cdots \quad \Rightarrow \quad \boxed{oldsymbol{x} = \Psi lpha}$$

- Recover (wavelet) coefficients α of image x.
- Consider the Laplacian prior on coefficients: $P(\alpha) \propto \exp\left(-\beta \|\alpha\|_{1}\right)$.
- Sparse synthesis regularisation problem:

$$oldsymbol{x}_{ ext{synthesis}} = oldsymbol{\Psi} imes rgmin_{oldsymbol{lpha}} \Big[oldsymbol{\|y - \Phi \Psi oldsymbol{lpha} ig\|_2^2 + \lambda oldsymbol{\|lpha\|}_1 \Big]$$

Synthesis framework

э

• Overcomplete dictionary composed of a concatenation of orthonormal bases:

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \dots, \boldsymbol{\Psi}_q \end{bmatrix}$$

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

$$oldsymbol{x}_{ ext{analysis}} = rgmin_{oldsymbol{x}} \Big[ig\| oldsymbol{y} - oldsymbol{\Phi} oldsymbol{x} ig\|_2^2 + \lambda ig\| oldsymbol{\Psi}^\dagger oldsymbol{x} ig\|_1 \Big]$$

Analysis framework

• Analysis problem viewed from a synthesis perspective:

$$\mathbf{x}_{\text{analysis}} = \mathbf{\Psi} \times \underset{\boldsymbol{\gamma} \in \text{column space}(\boldsymbol{\Psi}^{\dagger})}{\arg\min} \left[\left\| \boldsymbol{y} - \boldsymbol{\Phi} \boldsymbol{\Psi}^{\dagger} \boldsymbol{\gamma} \right\|_{2}^{2} + \lambda \left\| \boldsymbol{\gamma} \right\|_{1} \right]$$
Analysis as synthesis

• Overcomplete dictionary composed of a concatenation of orthonormal bases:

$$\mathbf{\Psi} = ig[\mathbf{\Psi}_1,\mathbf{\Psi}_2,\ldots,\mathbf{\Psi}_qig]$$

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

$$oldsymbol{x}_{ ext{analysis}} = rgmin_{oldsymbol{x}} \Big[ig\| oldsymbol{y} - oldsymbol{\Phi} oldsymbol{x} ig\|_2^2 + \lambda ig\| oldsymbol{\Psi}^\dagger oldsymbol{x} ig\|_1 \Big]$$

Analysis framework

・ロト ・ 日 ・ ・ ヨ ・ ・ ヨ ・

• Analysis problem viewed from a synthesis perspective:

$$\begin{aligned} \boldsymbol{x}_{\text{analysis}} = \boldsymbol{\Psi} \times & \arg\min_{\boldsymbol{\gamma} \in \mathsf{column space}(\boldsymbol{\Psi}^{\dagger})} \left[\left\| \boldsymbol{y} - \boldsymbol{\Phi} \boldsymbol{\Psi}^{\dagger} \boldsymbol{\gamma} \right\|_{2}^{2} + \lambda \left\| \boldsymbol{\gamma} \right\|_{1} \right] \\ & \text{Analysis as synthesis} \end{aligned}$$

• Overcomplete dictionary composed of a concatenation of orthonormal bases:

$$\mathbf{\Psi} = ig[\mathbf{\Psi}_1,\mathbf{\Psi}_2,\ldots,\mathbf{\Psi}_qig]$$

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

$$oldsymbol{x}_{ ext{analysis}} = rgmin_{oldsymbol{x}} \Big[ig\| oldsymbol{y} - oldsymbol{\Phi} oldsymbol{x} ig\|_2^2 + \lambda ig\| oldsymbol{\Psi}^\dagger oldsymbol{x} ig\|_1 \Big]$$

Analysis framework

• Analysis problem viewed from a synthesis perspective:

$$\begin{aligned} \boldsymbol{x}_{\text{analysis}} = \boldsymbol{\Psi} \times \min_{\boldsymbol{\gamma} \in \text{column space}(\boldsymbol{\Psi}^{\dagger})} \left[\left\| \boldsymbol{y} - \boldsymbol{\Phi} \boldsymbol{\Psi}^{\dagger} \boldsymbol{\gamma} \right\|_{2}^{2} + \lambda \left\| \boldsymbol{\gamma} \right\|_{1} \right] \\ \text{Analysis as synthesis} \end{aligned}$$

• Overcomplete dictionary composed of a concatenation of orthonormal bases:

$$\mathbf{\Psi} = ig[\mathbf{\Psi}_1,\mathbf{\Psi}_2,\dots,\mathbf{\Psi}_qig]$$

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

$$\mathbf{x}_{\text{analysis}} = \underset{\mathbf{x}}{\operatorname{arg\,min}} \left[\|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_{2}^{2} + \lambda \|\mathbf{\Psi}^{\dagger}\mathbf{x}\|_{1} \right]$$
Analysis framework

• Analysis problem viewed from a synthesis perspective:

$$\begin{aligned} \boldsymbol{x}_{\text{analysis}} = \boldsymbol{\Psi} \times \min_{\boldsymbol{\gamma} \in \mathsf{column space}(\boldsymbol{\Psi}^{\dagger})} \Big[\left\| \boldsymbol{y} - \boldsymbol{\Phi} \boldsymbol{\Psi}^{\dagger} \boldsymbol{\gamma} \right\|_{2}^{2} + \lambda \left\| \boldsymbol{\gamma} \right\|_{1} \Big] \\ \text{Analysis as synthesis} \end{aligned}$$
Compressive sensing More sophisticated approaches

• Overcomplete dictionary composed of a concatenation of orthonormal bases:

$$\mathbf{\Psi} = ig[\mathbf{\Psi}_1,\mathbf{\Psi}_2,\ldots,\mathbf{\Psi}_qig]$$

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

$$\boldsymbol{x}_{\text{analysis}} = \underset{\boldsymbol{x}}{\operatorname{arg\,min}} \left[\|\boldsymbol{y} - \boldsymbol{\Phi} \boldsymbol{x}\|_{2}^{2} + \lambda \|\boldsymbol{\Psi}^{\dagger} \boldsymbol{x}\|_{1} \right]$$
Analysis framework

• Analysis problem viewed from a synthesis perspective:

$$\begin{aligned} \boldsymbol{x}_{\text{analysis}} = \boldsymbol{\Psi} \times \min_{\boldsymbol{\gamma} \in \mathsf{column space}(\boldsymbol{\Psi}^{\dagger})} \Big[\left\| \boldsymbol{y} - \boldsymbol{\Phi} \boldsymbol{\Psi}^{\dagger} \boldsymbol{\gamma} \right\|_{2}^{2} + \lambda \left\| \boldsymbol{\gamma} \right\|_{1} \Big] \\ \text{Analysis as synthesis} \end{aligned}$$

イロト イポト イヨト イヨト

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

 $oldsymbol{x}^{i+1} \leftarrow \mathrm{P}(oldsymbol{x} \,|\, C^i_\ell, oldsymbol{y}) \quad \text{and} \quad C^{i+1}_\ell \leftarrow \mathrm{P}(C_\ell \,|\, oldsymbol{x}^{i+1}) \;.$

イロト イポト イヨト イヨト 二日

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

 $oldsymbol{x}^{i+1} \leftarrow \mathrm{P}(oldsymbol{x} \,|\, C^i_\ell, oldsymbol{y}) \quad \text{and} \quad C^{i+1}_\ell \leftarrow \mathrm{P}(C_\ell \,|\, oldsymbol{x}^{i+1}) \;.$

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

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

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

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

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

Outline





3 Sparse regularisation with Bayesian credible regions

イロト イロト イヨト イヨト 三日

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):

$$P(\boldsymbol{x} \mid \boldsymbol{y}) = \boxed{\pi(\boldsymbol{x})} \propto \exp\left[-\frac{g(\boldsymbol{x})}{Convex}\right]$$
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 π as stationary distribution:

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

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):

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

- 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:

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

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):

$$P(\boldsymbol{x} | \boldsymbol{y}) = \begin{bmatrix} \pi(\boldsymbol{x}) \\ Posterior \end{bmatrix} \propto \exp\left[-\begin{bmatrix} g(\boldsymbol{x}) \\ Convex \end{bmatrix}\right]$$

- 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:

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

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):

$$P(\boldsymbol{x} | \boldsymbol{y}) = \begin{bmatrix} \pi(\boldsymbol{x}) \\ Posterior \end{bmatrix} \propto \exp\left[-\begin{bmatrix} g(\boldsymbol{x}) \\ Convex \end{bmatrix}\right]$$

- 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:

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

where \mathcal{W} is Brownian motion.

Need gradients so cannot support sparse priors.

イロト イポト イヨト イヨト 二日

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):

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

- 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:

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

Gradient

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

Need gradients so cannot support sparse priors.

Proximity operators A brief aside

• Define proximity operator:

$$\mathrm{prox}_g^{\lambda}(\boldsymbol{x}) = \operatorname*{arg\,min}_{\boldsymbol{u}} \Big[g(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{x}\|^2 / 2\lambda \Big]$$

• Generalisation of projection operator:

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

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

・ロト ・ 日 ・ ・ ヨ ・ ・ 日 ・ ・

Proximity operators A brief aside

• Define proximity operator:

$$\operatorname{prox}_{g}^{\lambda}(\boldsymbol{x}) = \operatorname*{arg\,min}_{\boldsymbol{u}} \Big[g(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{x}\|^{2} / 2\lambda \Big]$$

• Generalisation of projection operator:

$$\mathcal{P}_{\mathcal{C}}(\boldsymbol{x}) = \operatorname*{arg\,min}_{\boldsymbol{u}} \Big[\imath_{\mathcal{C}}(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{x}\|^2/2 \Big],$$

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

Proximity operators A brief aside

• Define proximity operator:

$$\operatorname{prox}_{g}^{\lambda}(\boldsymbol{x}) = \operatorname*{arg\,min}_{\boldsymbol{u}} \Big[g(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{x}\|^{2} / 2\lambda \Big]$$

• Generalisation of projection operator:

$$\mathcal{P}_{\mathcal{C}}(\boldsymbol{x}) = \operatorname*{arg\,min}_{\boldsymbol{u}} \Big[\imath_{\mathcal{C}}(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{x}\|^2 / 2 \Big],$$

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 :

$$\pi_\lambda(oldsymbol{x}) = \sup_{oldsymbol{u} \in \mathbb{R}^N} \pi(oldsymbol{u}) \exp\Bigl(-rac{\|oldsymbol{u}-oldsymbol{x}\|^2}{2\lambda}\Bigr)$$

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



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

・ロト ・ 日 ・ ・ ヨ ・ ・ ヨ ・

Proximal MCMC Moreau approximation

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

$$\pi_\lambda(oldsymbol{x}) = \sup_{oldsymbol{u} \in \mathbb{R}^N} \pi(oldsymbol{u}) \exp\Bigl(-rac{\|oldsymbol{u}-oldsymbol{x}\|^2}{2\lambda}\Bigr)$$

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

$$As \ \lambda \to 0, \pi_{\lambda}(\boldsymbol{x}) \to \pi(\boldsymbol{x})$$



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

・ロト ・ 日 ・ ・ ヨ ・ ・ ヨ ・

• 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 δ controls the discrete-time increment and $w^{(m)}$ 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 $l^{(m)} o \pi$.)

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

• Discretise Langevin differential equation using forward Euler approximation:

$$\boldsymbol{l}^{(m+1)} = \boldsymbol{l}^{(m)} + \frac{\delta}{2} \boxed{\nabla \log \pi(\boldsymbol{l}^{(m)})} + \sqrt{\delta} \boldsymbol{w}^{(m)},$$
$$\nabla \log \pi_{\lambda}(\boldsymbol{x}) = (\operatorname{prox}_{g}^{\lambda}(\boldsymbol{x}) - \boldsymbol{x})/\lambda$$

where δ controls the discrete-time increment and $w^{(m)}$ 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 $l^{(m)} o \pi.)$
- Must compute $\mathrm{prox}_g^{\delta/2}(\cdot)$: develop analytic expression that can be computed rapidly.

(ロ) (四) (E) (E) (E) (E)

• Discretise Langevin differential equation using forward Euler approximation:

$$\boldsymbol{l}^{(m+1)} = \boldsymbol{l}^{(m)} + \frac{\delta}{2} \left[\nabla \log \pi(\boldsymbol{l}^{(m)}) + \sqrt{\delta} \boldsymbol{w}^{(m)}, \\ \nabla \log \pi_{\lambda}(\boldsymbol{x}) = (\operatorname{prox}_{g}^{\lambda}(\boldsymbol{x}) - \boldsymbol{x})/\lambda \right]$$

where δ controls the discrete-time increment and $w^{(m)}$ 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 $l^{(m)}
 ightarrow \pi$.)
- Must compute $\operatorname{prox}_{q}^{\delta/2}(\cdot)$: develop analytic expression that can be computed rapidly.

• Discretise Langevin differential equation using forward Euler approximation:

$$\boldsymbol{l}^{(m+1)} = \boldsymbol{l}^{(m)} + \frac{\delta}{2} \left[\nabla \log \pi(\boldsymbol{l}^{(m)}) + \sqrt{\delta} \boldsymbol{w}^{(m)}, \\ \nabla \log \pi_{\lambda}(\boldsymbol{x}) = (\operatorname{prox}_{g}^{\lambda}(\boldsymbol{x}) - \boldsymbol{x})/\lambda \right]$$

where δ controls the discrete-time increment and $w^{(m)}$ 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 $l^{(m)} o \pi$.)
- Must compute $\operatorname{prox}_g^{\delta/2}(\cdot)$: develop analytic expression that can be computed rapidly.

Proximal MCMC Preliminary results on simulations



(a) Dirty image

Figure: HII region of M31

Jason McEwen Statistical approaches for sparse imaging

ヘロト ヘアト ヘビト ヘビト

Proximal MCMC Preliminary results on simulations



(a) Dirty image

(b) Mean recovered image

Figure: HII region of M31

・ロト ・四ト ・ヨト ・ヨト

Proximal MCMC Preliminary results on simulations



Figure: HII region of M31

・ロト ・四ト ・ヨト ・ヨト

э

Proximal MCMC Preliminary results on simulations



(a) Dirty image

Figure: Cygnus A

Jason McEwen Statistical approaches for sparse imaging

ヘロト ヘアト ヘビト ヘビト

Proximal MCMC Preliminary results on simulations



(a) Dirty image

(b) Mean recovered image

Figure: Cygnus A

・ロト ・四ト ・ヨト ・ヨト

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

Figure: Supernova remnant W28

Jason McEwen Statistical approaches for sparse imaging

ヘロト ヘアト ヘビト ヘビト

Proximal MCMC Preliminary results on simulations



(a) Dirty image



Figure: Supernova remnant W28

イロト イロト イヨト イヨト

Proximal MCMC Preliminary results on simulations



Figure: Supernova remnant W28

・ロト ・日下・ ・日下・ ・日下

э

Proximal MCMC Preliminary results on simulations



(a) Dirty image

Figure: 3C288

ヘロト ヘアト ヘビト ヘビト

Proximal MCMC Preliminary results on simulations



(a) Dirty image

(b) Mean recovered image

Figure: 3C288

イロン イロン イヨン イヨン

Proximal MCMC Preliminary results on simulations



Figure: 3C288

イロト イロト イヨト イヨト

э

Outline







Sparse regularisation with Bayesian credible regions

▲□▶ ▲□▶ ★ □▶ ★ □▶ - □ - つくぐ

Bayesian credible regions Computation

- Combine error estimation with fast sparse regularisation (cf. compressive sensing) methods.
- Let C_{α} denote a Bayesian credible region with confidence level (1α) %:

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

• Define C_{α} by posterior iso-contour:

$$C_{lpha} := \{ oldsymbol{x} : g(oldsymbol{x}) \leq \gamma_{lpha} \} \; \;
ight |$$

• Analytic approximation of γ_{α} derived in Pereyra (2016b):

$$\tilde{\gamma}_{lpha} = g_{oldsymbol{y}}(oldsymbol{x}_{ ext{map}}) + N(au_{lpha}+1) ~,$$

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

• Compute $x_{
m map}$ by sparse regularisation and estimate Bayesian credible regions.

イロト 不得 トイヨト イヨト

Bayesian credible regions Computation

- Combine error estimation with fast sparse regularisation (cf. compressive sensing) methods.
- Let C_{α} denote a Bayesian credible region with confidence level (1α) %:

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

• Define C_{α} by posterior iso-contour:

$$C_{lpha} := \{ oldsymbol{x} : g(oldsymbol{x}) \leq \gamma_{lpha} \}$$

• Analytic approximation of γ_{α} derived in Pereyra (2016b):

$$\tilde{\gamma}_{lpha} = g_{oldsymbol{y}}(oldsymbol{x}_{ ext{map}}) + N(au_{lpha}+1) \ ,$$

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

• Compute $x_{
m map}$ 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_{α} denote a Bayesian credible region with confidence level (1α) %:

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

• Define C_{α} by posterior iso-contour:

$$C_{lpha} := \{ oldsymbol{x} : g(oldsymbol{x}) \leq \gamma_{lpha} \}$$

• Analytic approximation of γ_{α} derived in Pereyra (2016b):

$$\tilde{\gamma}_{lpha} = g_{oldsymbol{y}}(oldsymbol{x}_{ ext{map}}) + N(au_{lpha}+1) \ ,$$

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

• Compute $x_{
m map}$ 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_{α} denote a Bayesian credible region with confidence level $(1 \alpha)\%$:

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

• Define C_{α} by posterior iso-contour:

$$C_{lpha} := \{ oldsymbol{x} : g(oldsymbol{x}) \leq \gamma_{lpha} \}$$

• Analytic approximation of γ_{α} 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 $\alpha \in (4 \exp(-N/3), 1)$.

ullet Compute $x_{
m map}$ 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_{α} denote a Bayesian credible region with confidence level $(1 \alpha)\%$:

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

• Define C_{α} by posterior iso-contour:

$$C_{lpha} := \{ oldsymbol{x} : g(oldsymbol{x}) \leq \gamma_{lpha} \}$$

• Analytic approximation of γ_{α} 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 x_{map} by sparse regularisation and estimate Bayesian credible regions.

Bayesian Interpretations Proximal MCMC Bayesian Credibility

Bayesian credible regions Preliminary results on simulations



(a) Recovered image

Figure: HII region of M31

Jason McEwen Statistical approaches for sparse imaging

・ロト ・日下・ ・日下・ ・日下

Bayesian credible regions Preliminary results on simulations





- (a) Recovered image
- (b) Credible intervals for regions of size 10×10

Figure: HII region of M31

イロン イロン イヨン イヨン

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

э

イロン イロン イヨン イヨン

Figure: HII region of M31

Bayesian Interpretations Proximal MCMC Bayesian Credibility

Bayesian credible regions Preliminary results on simulations



(a) Recovered image

Figure: Cygnus A

Jason McEwen Statistical approaches for sparse imaging

・ロト ・日本・ ・ 日本・

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

イロン イロン イヨン イヨン

Figure: Cygnus A

Bayesian Interpretations Proximal MCMC Bayesian Credibility

Bayesian credible regions Preliminary results on simulations



(a) Recovered image

Figure: Supernova remnant W28

・ロト ・日下・ ・日下・ ・日下

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

イロト イポト イヨト イヨト

Figure: Supernova remnant W28

Bayesian Interpretations Proximal MCMC Bayesian Credibility

Bayesian credible regions Preliminary results on simulations



(a) Recovered image

Figure: 3C288

イロト イヨト イヨト イヨト

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

э

イロト イヨト イヨト イヨト

Figure: 3C288

• Is structure in an image physical or an artefact?

- Can we make precise statistical statements?
- Perform hypothesis tests using Bayesian credible regions (Pereyra 2016b).

・ロト ・聞ト ・ヨト ・ヨト

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

・ロト ・四ト ・ヨト ・ヨト

ъ.

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

イロト 不聞 と 不良 と 不良 とうほう

- 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

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

- 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

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

- 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

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

- 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

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

イロト 不得 トイヨト イヨト

- 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

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

イロト 不得 トイヨト イヨト



(a) Recovered image

Figure: HII region of M31

イロト イヨト イヨト イヨト



(a) Recovered image



Figure: HII region of M31

・ロト ・ 日 ・ ・ ヨ ・ ・ ヨ ・

э



(a) Recovered image

(b) Surrogate with region removed

Figure: HII region of M31



-0.2 -0.4 -0.6 -0.8

-1.6

イロト イポト イヨト イヨト

 $\Rightarrow \mathsf{structure} \ \mathsf{physical}$

Bayesian Interpretations Proximal MCMC Bayesian Credibility

Hypothesis testing Preliminary results on simulations



(a) Recovered image

Figure: Cygnus A

・ロト ・四ト ・ヨト ・ヨト



(a) Recovered image



(b) Surrogate with region removed

Figure: Cygnus A

Jason McEwen Statistical approaches for sparse imaging

イロン イロン イヨン イヨン



(a) Recovered image



(b) Surrogate with region removed

Figure: Cygnus A

Cannot reject null hypothesis

⇒ cannot make strong statistical statement about origin of structure

э

・ロト ・ 日 ・ ・ ヨ ・ ・ ヨ ・



(a) Recovered image

Figure: Supernova remnant W28

イロト イヨト イヨト イヨト



(a) Recovered image



Figure: Supernova remnant W28

イロト イポト イヨト イヨト



(a) Recovered image

(b) Surrogate with region removed

Figure: Supernova remnant W28

Reject null hypothesis ⇒ structure physical

イロト イポト イヨト イヨト



(a) Recovered image

Figure: 3C288

イロト イヨト イヨト イヨト



Figure: 3C288

Jason McEwen Statistical approaches for sparse imaging

・ロト ・ 日 ・ ・ ヨ ・ ・ ヨ ・



(a) Recovered image

- (b) Surrogate with region removed
 - Figure: 3C288

Reject null hypothesis ⇒ structure physical

42

-0.4 -0.6

4.8

-1.6

イロト イポト イヨト イヨト

- 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.

- 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.

- 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.

э

- 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.