Big Data in Cosmology

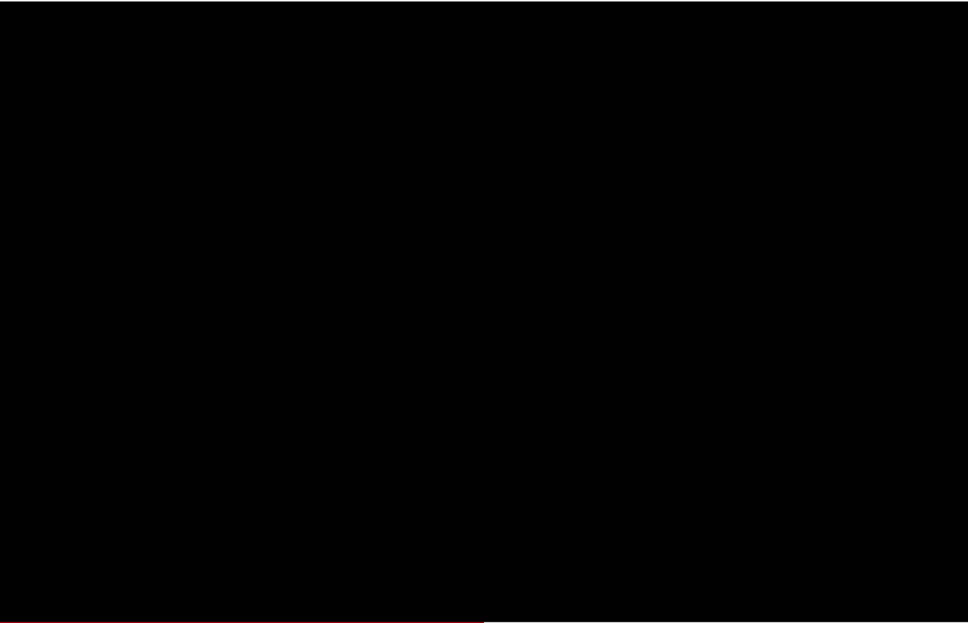
Data Intensive Science (DIS) in Cosmology

Jason McEwen
www.jasonmcewen.org
@jasonmcewen

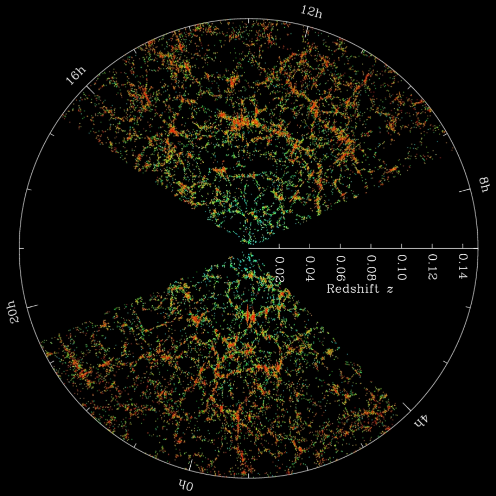
Mullard Space Science Laboratory (MSSL)
University College London (UCL)

Theory of Big Data Workshop
University College London (UCL), June 2017

Large-scale structure (LSS) of the Universe

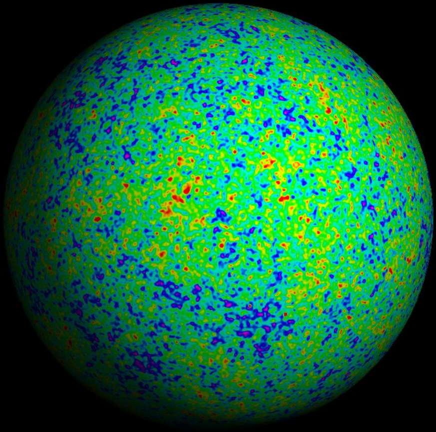


Observations of galaxies tracing large-scale structure (LSS)



Credit: SDSS

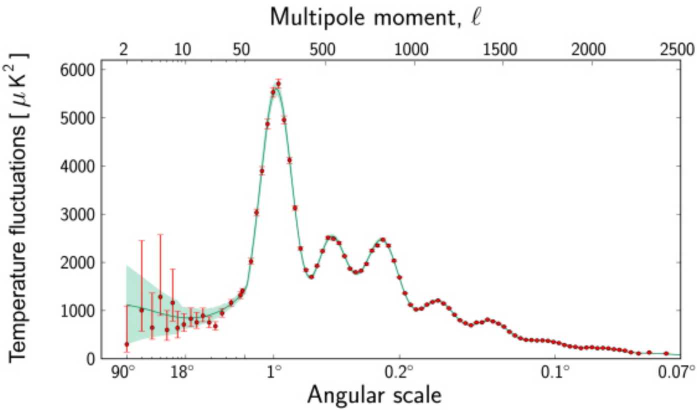
Observations of cosmic microwave background (CMB)



Credit: WMAP

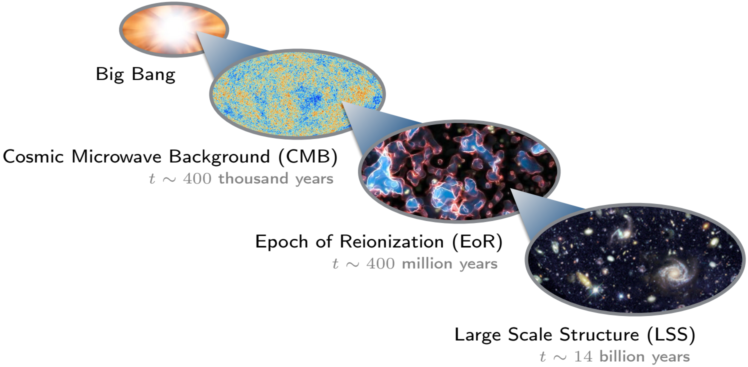
CMB power spectrum

Theory and observational data

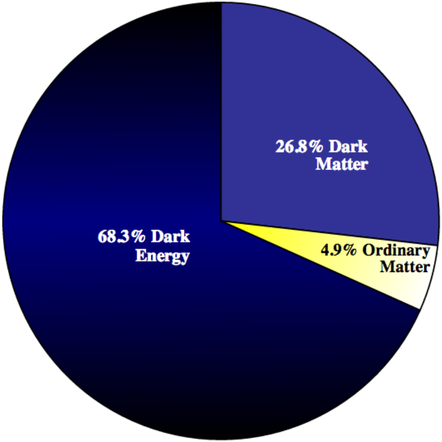


Credit: Planck

Cosmic evolution of our Universe

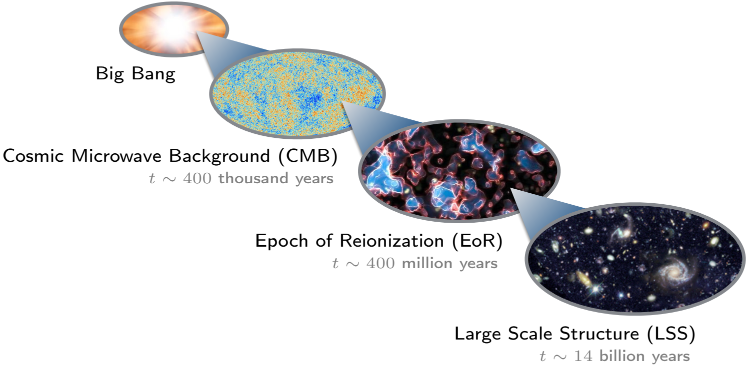


Content of the Universe

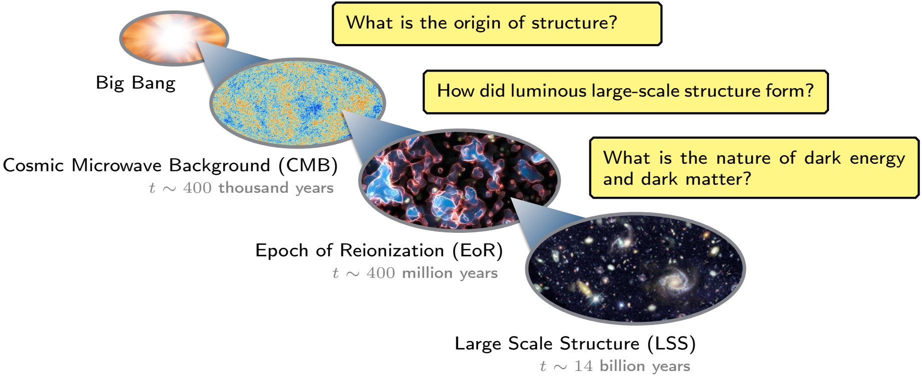


Credit: Planck

Unanswered fundamental questions



Unanswered fundamental questions

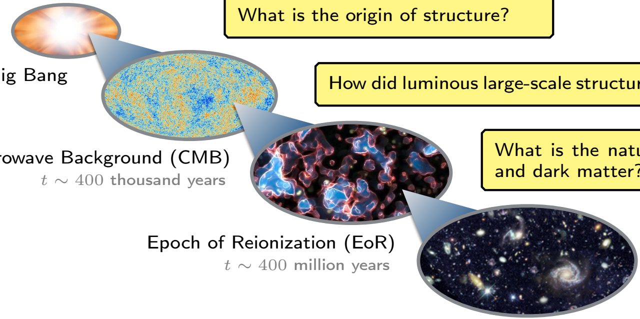


Unanswered fundamental questions

What is the origin of structure?

How did luminous large-scale structure form?

What is the nature of dark energy and dark matter?



Big Bang

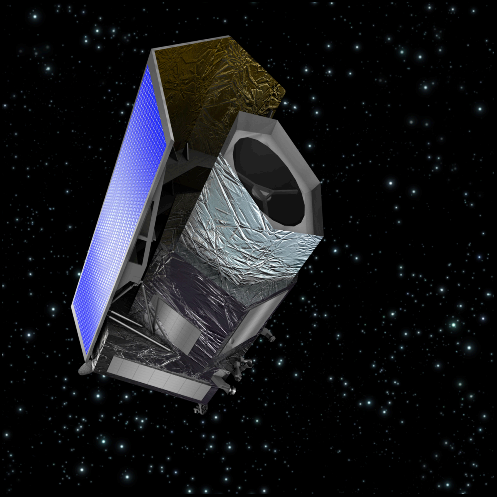
Cosmic Microwave Background (CMB)
$t \sim 400$ thousand years

Epoch of Reionization (EoR)
$t \sim 400$ million years

Large Scale Structure (LSS)
$t \sim 14$ billion years

Planck Gaia LOFAR SKA DES DESI Euclid LSST

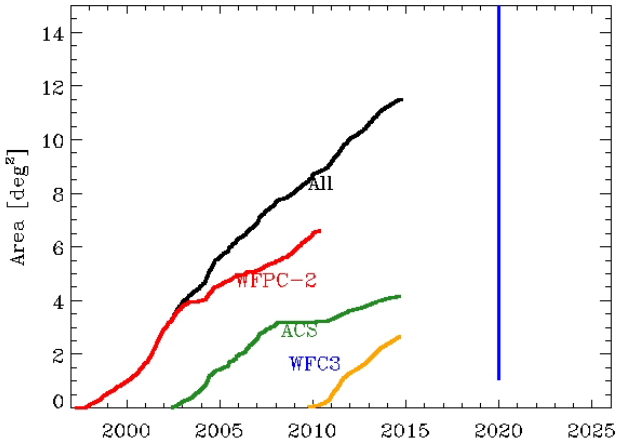
ESA Euclid satellite



Credit: Euclid

Euclid sky coverage

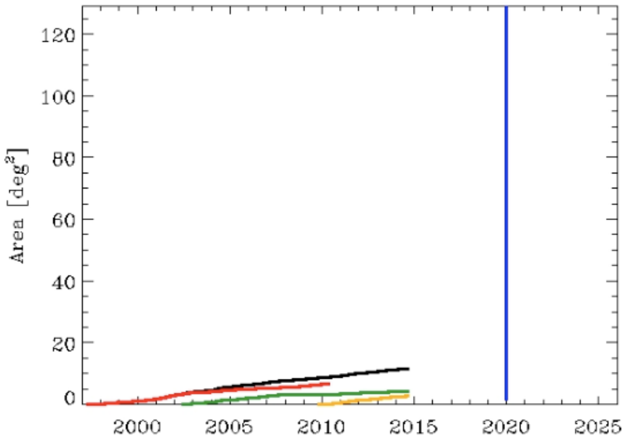
Switch on



Credit: Tom Kitching

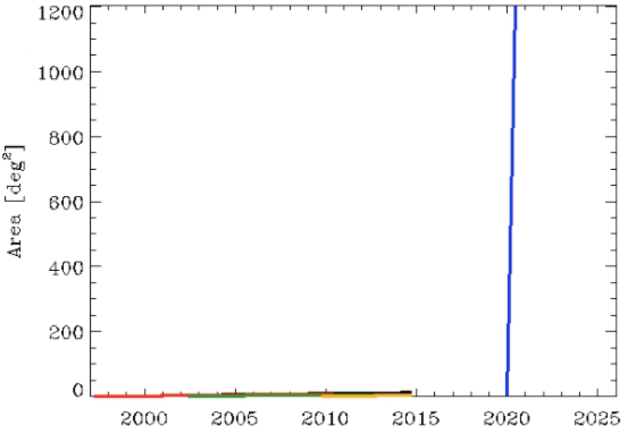
Euclid sky coverage

2 weeks



Credit: Tom Kitching

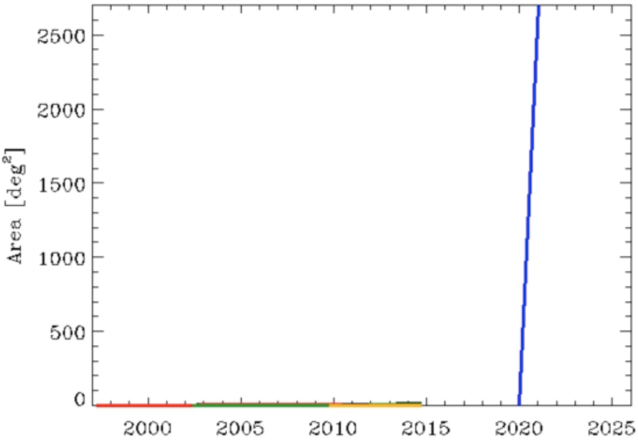
Euclid sky coverage
6 months



Credit: Tom Kitching

Euclid sky coverage

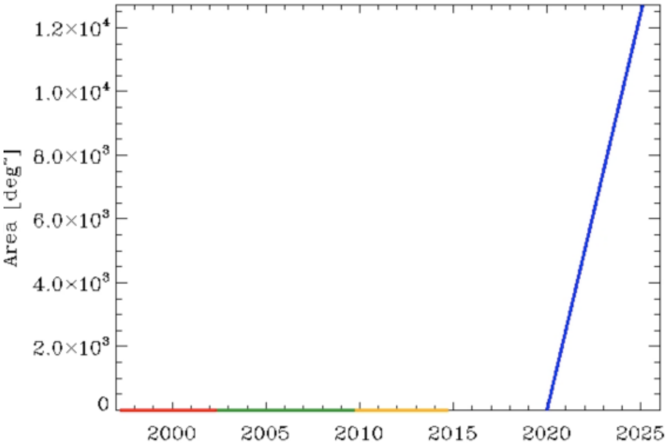
1 year



Credit: Tom Kitching

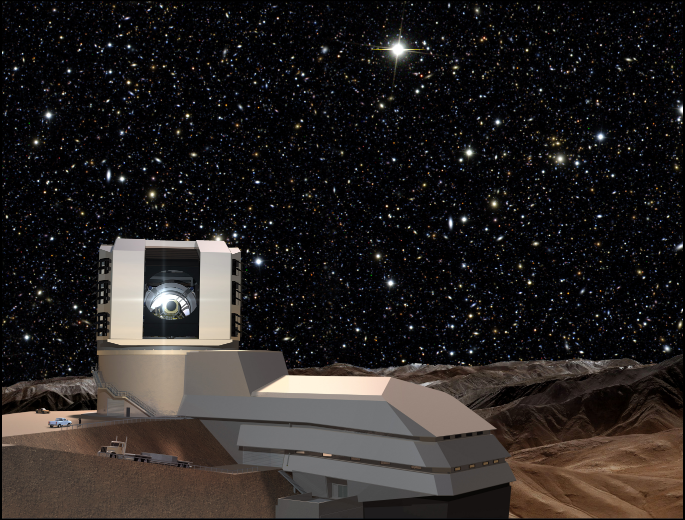
Euclid sky coverage

5 years



Credit: Tom Kitching

Large Synoptic Survey Telescope (LSST)



Credit: LSST

Large Synoptic Survey Telescope (LSST)

Data Releases:

Number of Data Releases = 11

Date of DR1 release = Date of Operations Start+ 12 months

Estimated numbers for DR-1 release

Objects = 18 billion

Sources = 350 billion (single epoch)

Forced Sources = 0.75 trillion

Estimated numbers for DR-11

Objects = 37 billion

Sources = 7 trillion (single epoch)

Forced Sources = 30 trillion

Visits observed = 2.75 million

Images collected = 5.5 million

Alert Production:

Real-time alert latency = 60 seconds

Average number of alerts per night= "about 10 million"

Data and compute sizes:

Final image collection (DR11) = 0.5 Exabytes

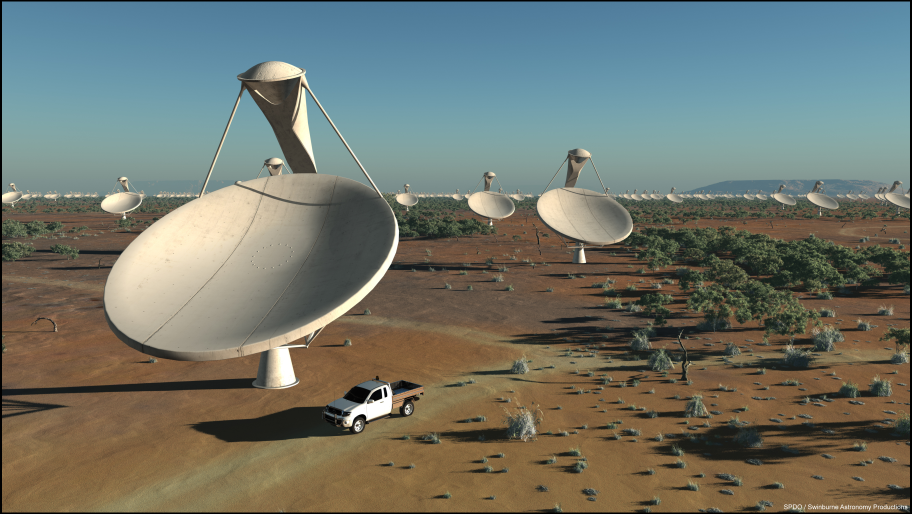
Final database size (DR11) = 15 PB

Final disk storage = 0.4 Exabytes

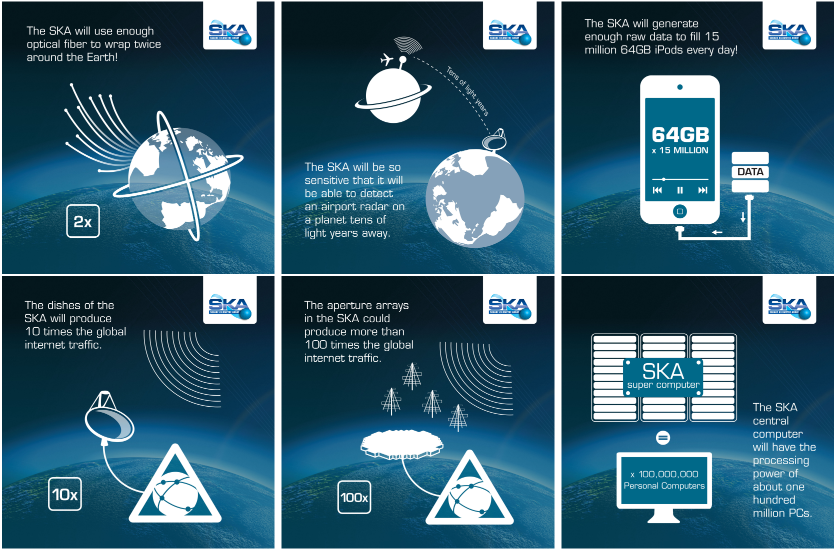
Peak number of nodes = 1750 nodes

Peak compute power in LSST data centers = 1.8 PFLOPS

Square Kilometre Array (SKA)



The SKA poses a considerable big-data challenge

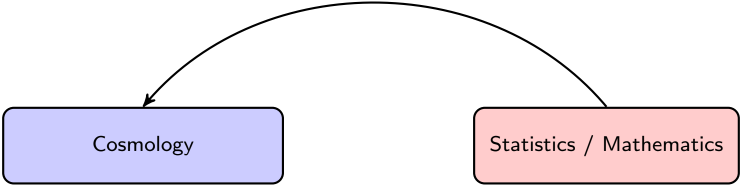


The SKA poses a considerable big-data challenge

Cosmostatistics & Cosmoinformatics

Closing the loop

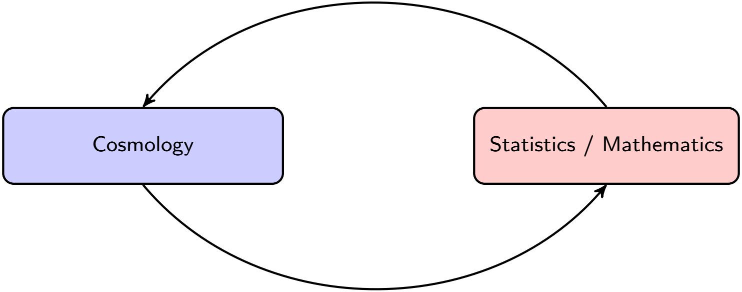
Extracting weak observational signatures of fundamental physics from complex data-sets requires sensitive, robust and principled analysis techniques.



Cosmostatistics & Cosmoinformatics

Closing the loop

Extracting weak observational signatures of fundamental physics from complex data-sets requires sensitive, robust and principled analysis techniques.



Constructing appropriate analysis techniques requires a deep understanding of cosmological problems and methodological foundations.

UCL Centre for Doctoral Training (CDT) in Data Intensive Science (DIS)

- UCL won bid to host STFC's first CDT.
  Learn more at out temporary website: https://www.hep.ucl.ac.uk/cdt-dis/

- Focused on Data Intensive Science (DIS).

- Aims:
  - Train next generation of leaders in the field of DIS (in both academic and industry).
  - Promote development and application of novel DIS techniques.
  - Promote knowledge transfer:
    - between academic fields;
    - between non-academic and academic organisations.

- Unique opportunity to bring together DIS research from perspective of applications, methodologies, and theoretical foundations.

UCL

Science & Technology Facilities Council

UCL Centre for Doctoral Training (CDT) in Data Intensive Science (DIS)

- UCL won bid to host STFC's first CDT.
  Learn more at out temporary website: https://www.hep.ucl.ac.uk/cdt-dis/
- Focused on Data Intensive Science (DIS).
- Aims:
  - Train next generation of leaders in the field of DIS (in both academic and industry).
  - Promote development and application of novel DIS techniques.
  - Promote knowledge transfer:
    - between academic fields;
    - between non-academic and academic organisations.
- Unique opportunity to bring together DIS research from perspective of applications, methodologies, and theoretical foundations.

UCL

Science & Technology Facilities Council

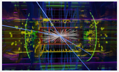
UCL Centre for Doctoral Training (CDT) in Data Intensive Science (DIS)

- UCL won bid to host STFC's first CDT.
  Learn more at out temporary website: https://www.hep.ucl.ac.uk/cdt-dis/
- Focused on Data Intensive Science (DIS).
- Aims:
  - Train next generation of leaders in the field of DIS (in both academic and industry).
  - Promote development and application of novel DIS techniques.
  - Promote knowledge transfer:
    - between academic fields;
    - between non-academic and academic organisations.
- Unique opportunity to bring together DIS research from perspective of applications, methodologies, and theoretical foundations.

UCL Centre for Doctoral Training (CDT) in Data Intensive Science (DIS)

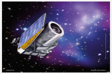
Who we are

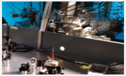
Particle Physics
Dpt. of Physics and Astronomy
(20 CDT Staff Members)



Astrophysics
Dpt. of Physics and Astronomy
(20 CDT Staff Members)

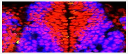
Department of
Space and Climate Science
(20 CDT Staff Members)

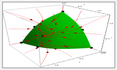




Atomic & Molecular Physics
Dpt. of Physics and Astronomy
(2 CDT Staff Members)

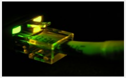
Department of
Computer Science
(8 CDT Staff Members)

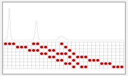




Department of
Mathematics
(9 CDT Staff Members)

Department of
Electrical Engineering
(3 CDT Staff Members)





Department of
Statistical Science
(5 CDT Staff Members)

Aim to foster closer collaboration between these areas to aid the development of novel DIS techniques or applications to new areas.

UCL Centre for Doctoral Training (CDT) in Data Intensive Science (DIS)

Management team

Centre Co-Directors: Profs N. Konstantinidis & O. Lahav


Directors of Research: Drs J. McEwen & T. Scanlon


Directors of Training: Prof. J. Tennyson FRS, & C. Gryce


Admissions & Graduate Tutor: Prof. S. Viti

Partner Liaison & Placements Co-Ordinator: Dr J. Yates


UCL Centre for Doctoral Training (CDT) in Data Intensive Science (DIS)

Industrial partners

IT Companies

Data Intensive Companies

Public Sector Organisations

Non-Academic Research

Public-Private Partnerships

- Students will undertake 6 month internships with partners on a DIS project
- Promote knowledge transfer between academic and non-academic organisations.
- We've been approached by more organisations since winning the bid (UKAEA, Asos, GroupM, S&P, Illuminas, ASI, . . .).

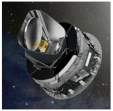
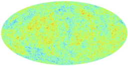
Outline

1. Bayesian inference
2. Computational harmonic analysis
3. Machine learning
4. Inverse problems

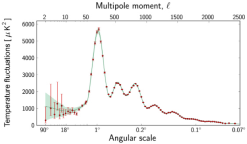
Outline

1. Bayesian inference
2. Computational harmonic analysis
3. Machine learning
4. Inverse problems

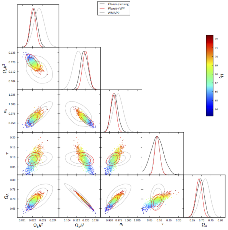
Bayesian inference for parameter estimation
Case study: CMB

Observation



$$P(\Theta|d, M) \propto P(d|\Theta, M)P(\Theta|M)$$



Bayesian Inference

$$\Theta^\star = (A_s^\star, n_s^\star, \tau^\star, \Omega_\mathrm{b}^\star h^2, \Omega_\mathrm{c}^\star h^2, \Omega_\Lambda^\star)$$

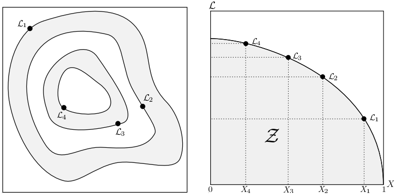
$$\Theta = (A_s, n_s, \tau, \Omega_\mathrm{b} h^2, \Omega_\mathrm{c} h^2, \Omega_\Lambda)$$

Theory

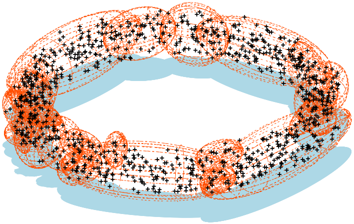
Figure: CMB Bayesian inference pipeline.

Bayesian inference for model selection

- Nested sampling (Skilling 2005).
- MultiNest: multi-modal ellipsoidal sampling (Feroz & Hobson 2007; Feroz, Hobson & Bridges 2008).
- PolyChord: multi-modal whitened slice sampling (Handley, Hobson & Lasenby 2015).



(a) Nested sampling



(b) Ellipsoidal sampling

Figure: Computing the marginalised likelihood (Bayesian evidence) [Credit: Feroz *et al.* 2008].

Bayesian hierarchical models

Weak gravitational lensing

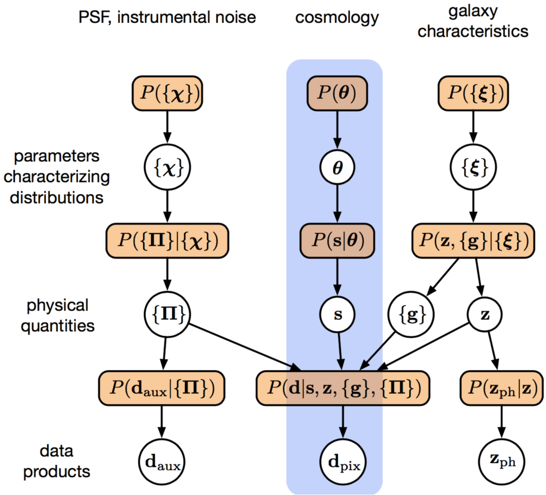
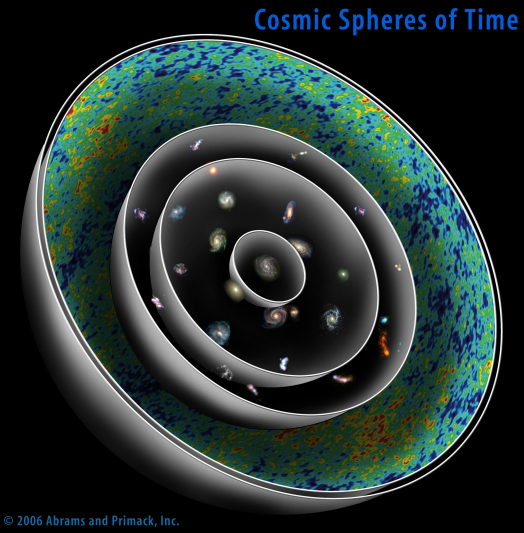


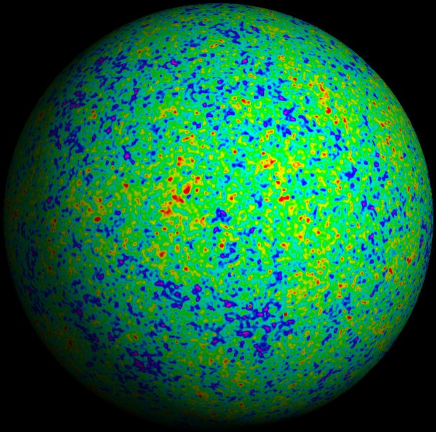
Figure: HBM for weak gravitational lensing (Alsing *et al.* 2015)

Outline

Observations made on the celestial sphere



Cosmic microwave background (CMB) on the celestial sphere

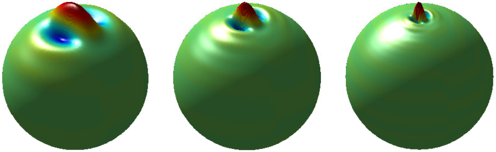


Credit: WMAP

Wavelets on the sphere

- Spin scale-discretised wavelet transform given by projection onto each wavelet (McEwen *et al.* 2015; McEwen 2015; McEwen *et al.* 2013; Wiaux, McEwen *et al.* 2008):

$$\underbrace{\boxed{W^{{}_s\Psi^j}(\rho) = \langle {}_s f, \mathcal{R}_\rho\, {}_s\Psi^j \rangle}}_{\text{projection}} = \int_{\mathbb{S}^2} \mathrm{d}\Omega(\theta,\varphi)\, {}_s f(\theta,\varphi)\, (\mathcal{R}_\rho\, {}_s\Psi^j)^*(\theta,\varphi)\ .$$



(a) $j = 3$ (b) $j = 4$ (c) $j = 5$

Figure: Wavelets on sphere

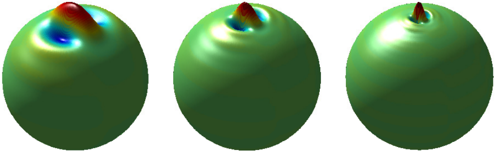
- Original function may be recovered exactly in practice from wavelet coefficients:

$${}_s f(\omega) = \underbrace{\boxed{\sum_{j=0}^{J}}}_{\text{finite sum}} \underbrace{\boxed{\int_{\mathrm{SO}(3)} \mathrm{d}\varrho(\rho)\, W^{{}_s\Psi^j}(\rho)\, (\mathcal{R}_\rho\, {}_s\Psi^j)(\omega)}}_{\text{wavelet contribution}}\ .$$

Wavelets on the sphere

- Spin scale-discretised wavelet transform given by projection onto each wavelet (McEwen *et al.* 2015; McEwen 2015; McEwen *et al.* 2013; Wiaux, McEwen *et al.* 2008):

$$\underbrace{\boxed{W^{{}_s\Psi^j}(\rho) = \langle {}_sf, \mathcal{R}_\rho\, {}_s\Psi^j \rangle}}_{\text{projection}} = \int_{\mathbb{S}^2} \mathrm{d}\Omega(\theta,\varphi)\; {}_sf(\theta,\varphi)\, (\mathcal{R}_\rho\, {}_s\Psi^j)^*(\theta,\varphi)\,.$$



(a) $j = 3$ (b) $j = 4$ (c) $j = 5$

Figure: Wavelets on sphere

- Original function may be recovered exactly in practice from wavelet coefficients:

$${}_sf(\omega) = \underbrace{\boxed{\sum_{j=0}^{J}}}_{\text{finite sum}} \underbrace{\boxed{\int_{\mathrm{SO}(3)} \mathrm{d}\varrho(\rho)\, W^{{}_s\Psi^j}(\rho)\, (\mathcal{R}_\rho\, {}_s\Psi^j)(\omega)}}_{\text{wavelet contribution}}\,.$$

Wavelets on the sphere

Localisation of Gaussian random fields

Wavelet localisation (McEwen *et al.* 2016)

Directional scale-discretised wavelets $\Psi \in \mathrm{L}^2(\mathbb{S}^2)$, defined on the sphere $\mathbb{S}^2$ and centred on the North pole, satisfy the localisation bound:

$$\boxed{\left|\Psi^{(j)}(\theta,\varphi)\right| \leq \frac{C_1^{(j)}}{\left(1 + C_2^{(j)}\,\theta\right)^{\xi}}}$$

(there exist strictly positive constants $C_1^{(j)}, C_2^{(j)} \in \mathbb{R}_*^+$ for any $\xi \in \mathbb{R}_*^+$).

Wavelet asymptotic uncorrelation (McEwen *et al.* 2016)

For Gaussian random fields on the sphere, directional scale-discretised wavelet coefficients are asymptotically uncorrelated. The directional wavelet correlation satisfies the bound:

$$\boxed{\Xi^{(jj')}(\rho_1,\rho_2) \leq \frac{C_1^{(j)}}{\left(1 + C_2^{(j)}\beta\right)^{\xi}}},$$

where $\beta \in [0,\pi)$ is an angular separation between Euler angles ρ_1 and ρ_2 (there exist strictly positive constants $C_1^{(j)}, C_2^{(j)} \in \mathbb{R}_*^+$ for any $\xi \in \mathbb{R}_*^+$, $\xi \geq 2M$, where M is the azimuthal band-limit of the wavelet and $\left|j - j'\right| < 2$).

Wavelets on the sphere

Localisation of Gaussian random fields

Wavelet localisation (McEwen *et al.* 2016)

Directional scale-discretised wavelets $\Psi \in \mathrm{L}^2(\mathbb{S}^2)$, defined on the sphere $\mathbb{S}^2$ and centred on the North pole, satisfy the localisation bound:

$$\boxed{\left|\Psi^{(j)}(\theta,\varphi)\right| \leq \frac{C_1^{(j)}}{\left(1 + C_2^{(j)}\,\theta\right)^{\xi}}}$$

(there exist strictly positive constants $C_1^{(j)}, C_2^{(j)} \in \mathbb{R}_*^+$ for any $\xi \in \mathbb{R}_*^+$).

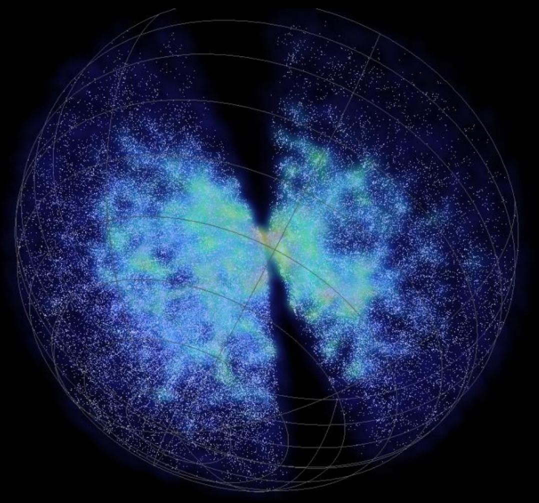
Wavelet asymptotic uncorrelation (McEwen *et al.* 2016)

For Gaussian random fields on the sphere, directional scale-discretised wavelet coefficients are asymptotically uncorrelated. The directional wavelet correlation satisfies the bound:

$$\boxed{\Xi^{(jj')}(\rho_1,\rho_2) \leq \frac{C_1^{(j)}}{\left(1 + C_2^{(j)}\beta\right)^{\xi}}}\,,$$

where $\beta \in [0,\pi)$ is an angular separation between Euler angles ρ_1 and ρ_2 (there exist strictly positive constants $C_1^{(j)}, C_2^{(j)} \in \mathbb{R}_*^+$ for any $\xi \in \mathbb{R}_*^+$, $\xi \geq 2M$, where M is the azimuthal band-limit of the wavelet and $\left|j - j'\right| < 2$).

Galaxy distribution tracing large-scale structure (LSS) on the 3D ball



Credit: SDSS

Wavelets on the ball

Fourier-LAGuerre wavelets (flaglets)

- Fourier-Laguerre wavelet (flaglet) transform is given by the projection onto each wavelet (Leistedt & McEwen 2012):

$$\underbrace{\boxed{W^{{}_s\Psi^{jj'}}(r,\rho) = \langle {}_s f,\ \mathcal{T}_{(r,\rho)}\ {}_s\Psi^{jj'} \rangle}}_{\text{projection}} = \int_{\mathbb{B}^3} \mathrm{d}^3\boldsymbol{r}\ {}_s f(\boldsymbol{r}) (\mathcal{T}_{(r,\rho)}\ {}_s\Psi^{jj'})^*(\boldsymbol{r})\ .$$

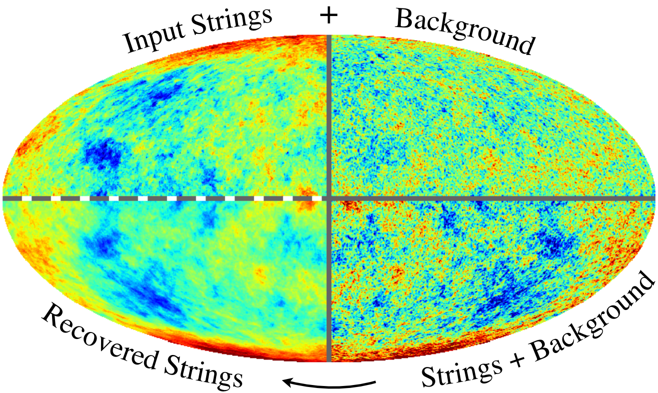
- Original function may be recovered exactly in practice from wavelet coefficients:

$${}_s f(\boldsymbol{r}) = \underbrace{\boxed{\sum_{j\,j'}}}_{\text{finite sum}} \underbrace{\boxed{\int_{\mathrm{SO}(3)} \mathrm{d}\varrho(\rho) \int_{\mathbb{R}+} \mathrm{d}r\, W^{{}_s\Psi^{jj'}}(r,\rho) (\mathcal{T}_{(r,\rho)}\ {}_s\Psi^{jj'})(\boldsymbol{r})}}_{\text{wavelet contribution}}\ .$$

- Opens up wavelet analyses of galaxy distribution tracing the large-scale structure (LSS).

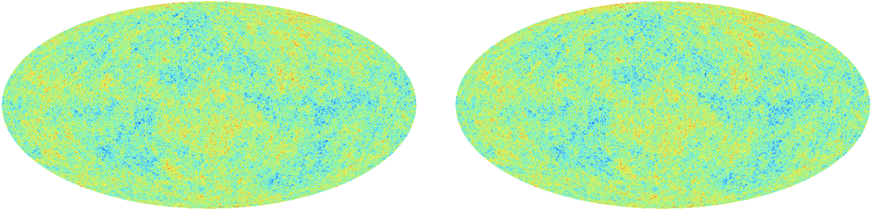
Cosmic strings

Problem



Cosmic strings

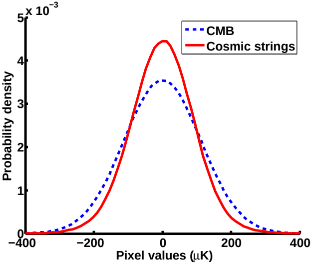
Typical amplitude



(a) CMB (b) CMB with embedded string contribution

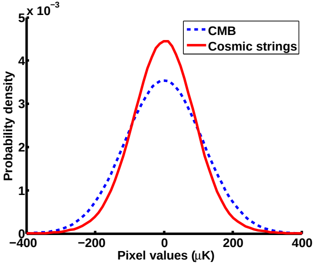
Figure: CMB simulation with string contribution embedded ($G\mu = 5 \times 10^{-7}$).

Cosmic strings
Wavelet representation

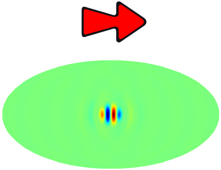


Cosmic strings

Wavelet representation

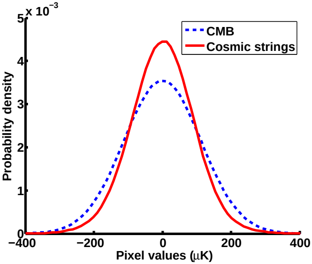


Wavelet transform

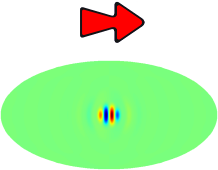


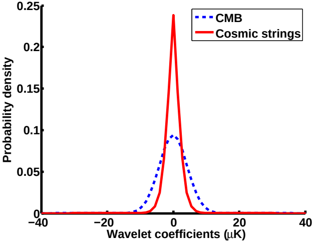
Cosmic strings

Wavelet representation



Wavelet transform





Cosmic strings

Hierarchical Bayesian model

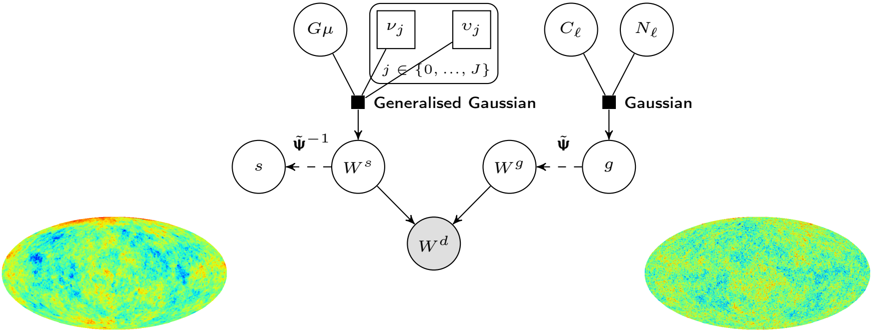


Figure: Hierarchical Bayesian model (McEwen *et al.* 2016)

Cosmic strings
Bayesian inference

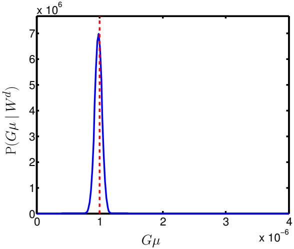


Figure: Posterior

Table: Bayes factors

$G\mu$ truth / 10^{-7}	Bayes factor [$\log_e$]
10.0	51.4
7.00	12.5
5.00	1.19
3.00	-3.87

Cosmic strings
Bayesian inference

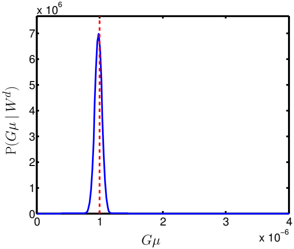
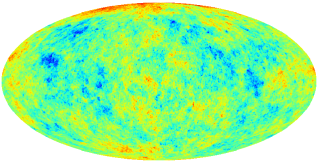


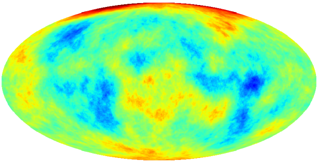
Figure: Posterior

Table: Bayes factors

$G\mu$ truth / 10^{-7}	Bayes factor $[\log_e]$
10.0	51.4
7.00	12.5
5.00	1.19
3.00	−3.87



(a) Ground truth

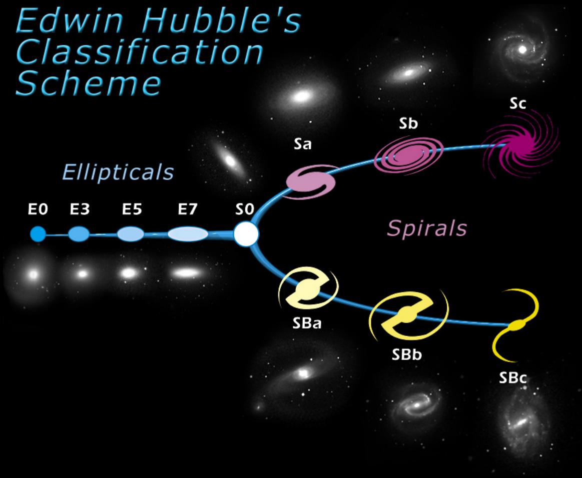


(b) Recovered

Figure: String map

Outline

Galaxy morphology classification



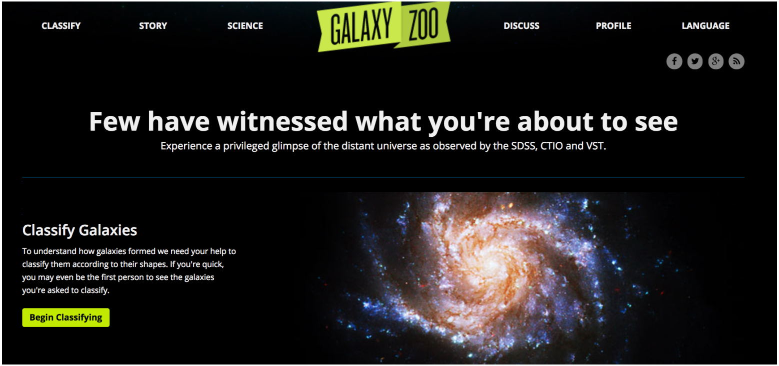
Credit: Wikipedia

Galaxy morphology classification

- Galaxy classification with neural networks pioneered by Lahav in 1990s (Lahav, Naim *et al.* 1995; Banerji, Lahav *et al.* 2009).
- Galaxy Zoo to crowdsource galaxy classification → ~50 million classifications / year.
- For upcoming surveys with ~1.5 billion galaxies, would take 30 years!

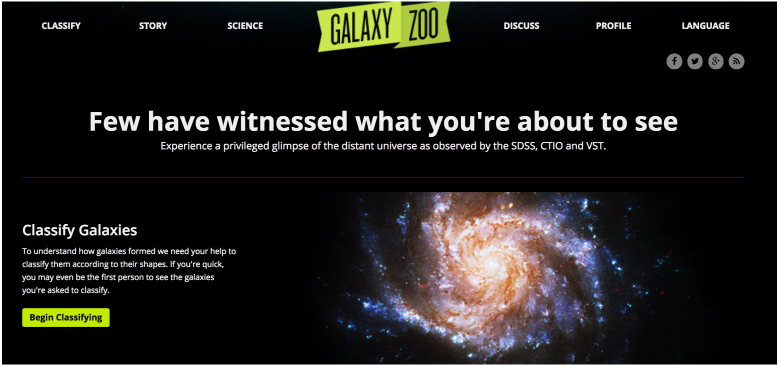
Galaxy morphology classification

- Galaxy classification with neural networks pioneered by Lahav in 1990s (Lahav, Naim *et al.* 1995; Banerji, Lahav *et al.* 2009).
- Galaxy Zoo to crowdsource galaxy classification → ~50 million classifications / year.
- For upcoming surveys with ~1.5 billion galaxies, would take 30 years!



Galaxy morphology classification

- Galaxy classification with neural networks pioneered by Lahav in 1990s (Lahav, Naim *et al.* 1995; Banerji, Lahav *et al.* 2009).
- Galaxy Zoo to crowdsource galaxy classification $\rightarrow$ $\sim$50 million classifications / year.
- For upcoming surveys with $\sim$1.5 billion galaxies, would take 30 years!



Galaxy morphology classification

- Use Galazy Zoo classification as training data (Lahav, Olhede, *et al.*, ongoing).

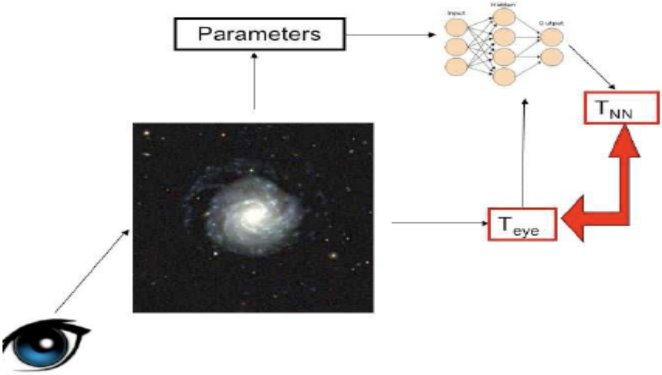


Figure: Crowdsourcing and machine learning for galaxy classification **[Credit: Lahav]**

Photometric redshift estimation

- Photometric redshift estimation with neural networks pioneered by Lahav in 2000s (Collister & Lahav 2004; Sadeh, Abdalla & Lahav 2016).

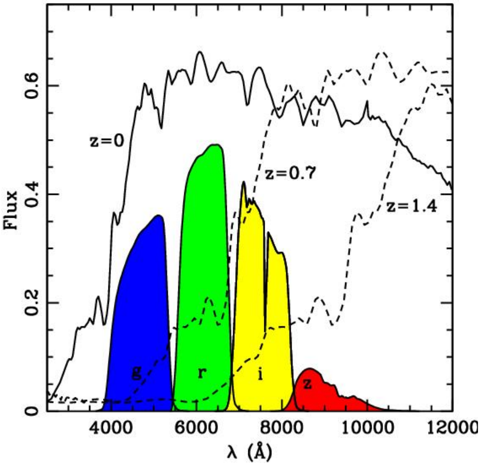


Figure: Photometric redshift estimation **[Credit: Lahav]**

Artist impression of Supernova explosion

Thermonuclear explosion or core collapse

Supernova classification

Spectroscopic classification

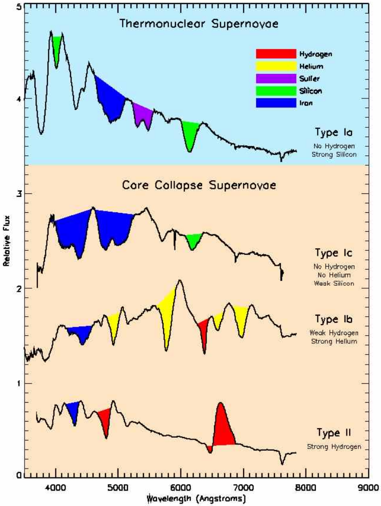


Figure: Spectroscopic observations

Supernova classification

Photometric classification

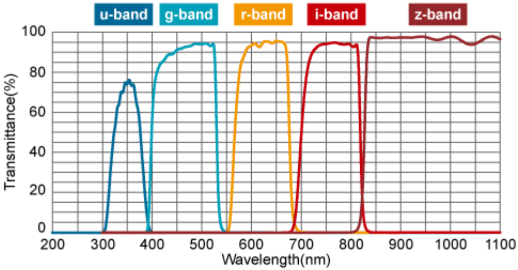
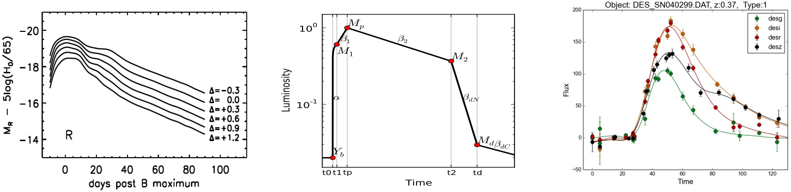


Figure: Photometric observations.

Supernova classification

Photometric classification

- Photometric Supernova classification by machine learning (Lochner, McEwen, Peiris, Lahav & Winter 2016)
- Go beyond single techniques to study classes.



(a) Templates (b) Generic parameterisations (c) Wavelets (non-parametric)

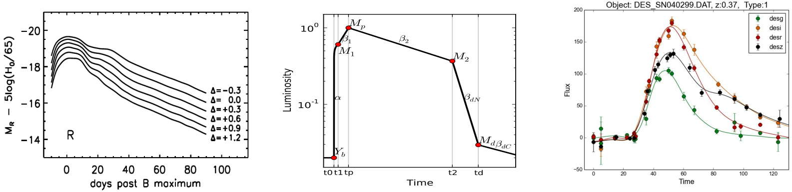
Figure: Feature selection classes (in order of increasing model independence)

- Integrate physics into machine learning (scale and dilation invariance).

Supernova classification

Photometric classification

- Photometric Supernova classification by machine learning (Lochner, McEwen, Peiris, Lahav & Winter 2016)
- Go beyond single techniques to study classes.



(a) Templates (b) Generic parameterisations (c) Wavelets (non-parametric)

Figure: Feature selection classes (in order of increasing model independence)

- Integrate physics into machine learning (scale and dilation invariance).

Supernova classification

Representativeness of training data

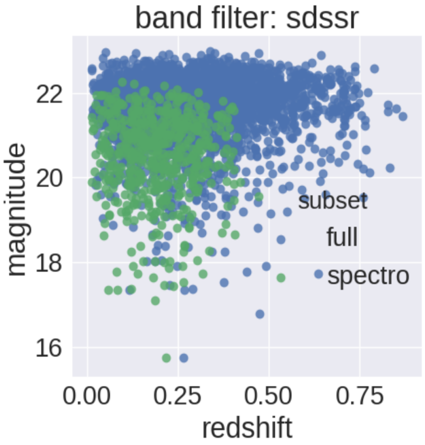
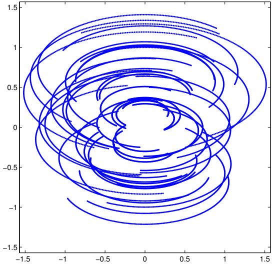


Figure: Training (green) vs test (blue) data

Outline

Radio interferometric telescopes acquire “Fourier” measurements

“Fourier” Measurements



Radio interferometric inverse problem

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

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

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

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

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

Radio interferometric inverse problem

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

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

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

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

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

Radio interferometric inverse problem

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

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

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

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

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

Sparse regularisation

Synthesis and analysis frameworks

- Sparse synthesis regularisation problem:

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

Synthesis framework

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

- Sparse analysis regularisation problem (Elad *et al.* 2007, Nam *et al.* 2012):

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

Analysis framework

- For orthogonal bases the two approaches are identical but otherwise very different.
- Sparsity averaging reweighted analysis (SARA)
(Carrillo, McEwen & Wiaux 2012; Carrillo, McEwen, Van De Ville, Thiran & Wiaux 2013).

Sparse regularisation

Synthesis and analysis frameworks

- Sparse synthesis regularisation problem:

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

Synthesis framework

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

- Sparse analysis regularisation problem (Elad *et al.* 2007, Nam *et al.* 2012):

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

Analysis framework

- For orthogonal bases the two approaches are identical but otherwise very different.
- Sparsity averaging reweighted analysis (SARA) (Carrillo, McEwen & Wiaux 2012; Carrillo, McEwen, Van De Ville, Thiran & Wiaux 2013).

Sparse regularisation

Synthesis and analysis frameworks

- Sparse synthesis regularisation problem:

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

Synthesis framework

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

- Sparse analysis regularisation problem (Elad *et al.* 2007, Nam *et al.* 2012):

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

Analysis framework

- For orthogonal bases the two approaches are identical but otherwise very different.
- Sparsity averaging reweighted analysis (SARA)
(Carrillo, McEwen & Wiaux 2012; Carrillo, McEwen, Van De Ville, Thiran & Wiaux 2013).

Public open-source codes

PURIFY code

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

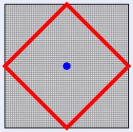
Next-generation radio interferometric imaging

Carrillo, McEwen, Wiaux, Pratley, d'Avezac

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

SOPT code

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



Sparse OPTimisation

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

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

Imaging observations from the VLA and ATCA with PURIFY

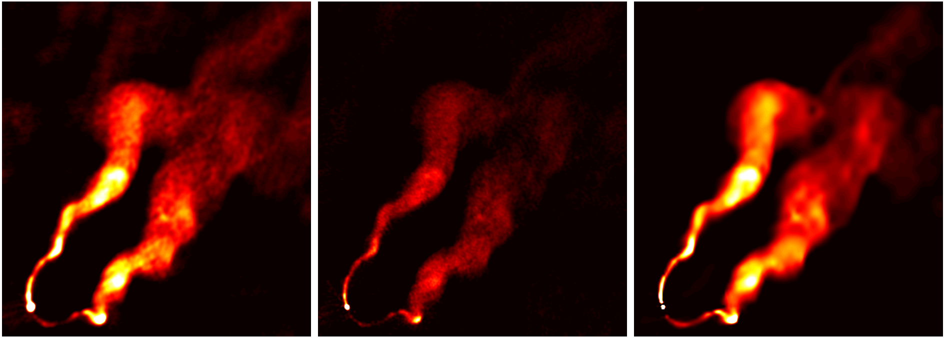
(a) NRAO Very Large Array (VLA)

(b) Australia Telescope Compact Array (ATCA)

Figure: Radio interferometric telescopes considered

PURIFY reconstruction

VLA observation of 3C129

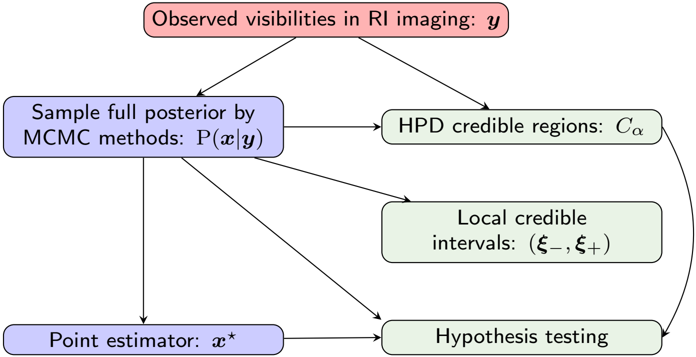


(a) CLEAN (natural) (b) CLEAN (uniform) (c) PURIFY

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

Proximal MCMC sampling and uncertainty quantification

- See poster by Xiaohao Cai
(Cai, Pereyra & McEwen, 2017a, in prep.; Cai, Pereyra & McEwen, 2017b, in prep.)



Sampling the full posterior distribution

Markov Chain Monte Carlo (MCMC)

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

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

Sampling the full posterior distribution

Markov Chain Monte Carlo (MCMC)

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

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

Sampling the full posterior distribution

Markov Chain Monte Carlo (MCMC)

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

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

MCMC sampling with gradients

Langevin dynamics

- Consider posteriors of the following form (and more compact notation):

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

- If $g(x)$ differentiable can adopt MALA (Langevin dynamics) or HMC (Hamiltonian dynamics) MCMC methods.
- MALA 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):

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

- If $g(\boldsymbol{x})$ differentiable can adopt MALA (Langevin dynamics) or HMC (Hamiltonian dynamics) MCMC methods.

- MALA 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):

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

- If $g(\boldsymbol{x})$ differentiable can adopt MALA (Langevin dynamics) or HMC (Hamiltonian dynamics) MCMC methods.

- MALA 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):

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

- If $g(\boldsymbol{x})$ differentiable can adopt MALA (Langevin dynamics) or HMC (Hamiltonian dynamics) MCMC methods.

- MALA based on Langevin diffusion process $\mathcal{L}(t)$, with π as stationary distribution:

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

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

- Need gradients so cannot support sparse priors.

Proximity operators

A brief aside

- Define proximity operator:

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

- 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.

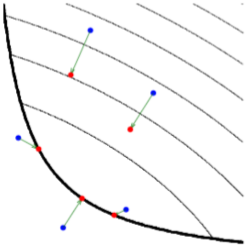


Figure: Illustration of proximity operator [Credit: Parikh & Boyd (2013)]

Proximal MALA

MCMC sampling

Proximal Metropolis adjusted Langevin algorithm (P-MALA)

Pereyra (2016a)

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

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

- Euler discretisation and apply Moreau approximation to π:

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

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

- Metropolis-Hastings accept-reject step.

Proximal MALA

MCMC sampling

Proximal Metropolis adjusted Langevin algorithm (P-MALA)
Pereyra (2016a)

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

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

- Euler discretisation and apply Moreau approximation to π:

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

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

- Metropolis-Hastings accept-reject step.

Proximal MALA

MCMC sampling

Proximal Metropolis adjusted Langevin algorithm (P-MALA)

Pereyra (2016a)

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

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

- Euler discretisation and apply Moreau approximation to π:

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

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

- Metropolis-Hastings accept-reject step.

Proximal MALA

MCMC sampling

Proximal Metropolis adjusted Langevin algorithm (P-MALA)

Pereyra (2016a)

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

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

- Euler discretisation and apply Moreau approximation to π:

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

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

- Metropolis-Hastings accept-reject step.

Proximal MALA

Computing proximity operators for the analysis case

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

  Prior Likelihood

- Must solve an optimisation problem for each iteration!

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

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

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

  Single forward-backward iteration

- Analytic approximation:

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

Proximal MALA

Computing proximity operators for the analysis case

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

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

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

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

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

  Single forward-backward iteration
- Analytic approximation:

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

Proximal MALA

Computing proximity operators for the analysis case

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

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

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

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

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

Single forward-backward iteration

- Analytic approximation:

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

Proximal MALA

Computing proximity operators for the analysis case

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

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

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

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

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

Single forward-backward iteration

- Analytic approximation:

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

MYULA
MCMC sampling

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

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

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

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

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

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

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

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

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

MYULA

MCMC sampling

Moreau-Yosida unadjusted Langevin algorithm (MYULA)

Durmus, Moulines & Pereyra (2016)

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

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

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

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

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

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

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

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

MYULA

MCMC sampling

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

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

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

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

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

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

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

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

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

MYULA
MCMC sampling

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

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

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

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

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

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

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

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

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

MYULA

Computing proximity operators for the analysis case

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

  Prior Likelihood
- Only need to compute proximity operator of $\bar{f}_1$, which can be computed analytically without any approximation:

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

MYULA

Computing proximity operators for the analysis case

- Recall posterior: $\pi(\boldsymbol{x}) \propto \exp\big(-g(\boldsymbol{x})\big)$.

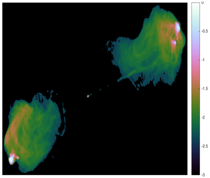
- Let $\bar{g}(\boldsymbol{x}) = \bar{f}_1(\boldsymbol{x}) + \bar{f}_2(\boldsymbol{x})$, where $\underbrace{\boxed{\bar{f}_1(\boldsymbol{x}) = \mu \|\boldsymbol{\Psi}^\dagger \boldsymbol{x}\|_1}}_{\text{Prior}}$ and $\underbrace{\boxed{\bar{f}_2(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 / 2\sigma^2}}_{\text{Likelihood}}$.

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

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

Numerical experiments

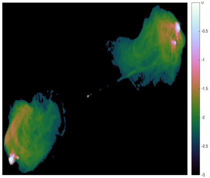
MYULA with analysis model



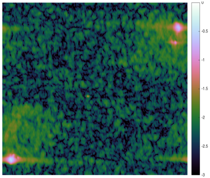
(a) Ground truth

Figure: Cygnus A

Numerical experiments
MYULA with analysis model



(a) Ground truth

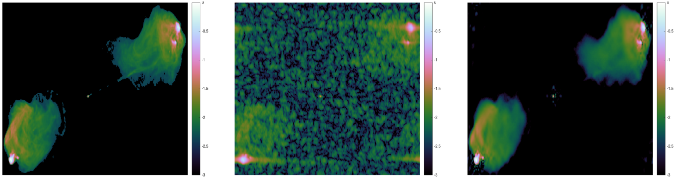


(b) Dirty image

Figure: Cygnus A

Numerical experiments

MYULA with analysis model

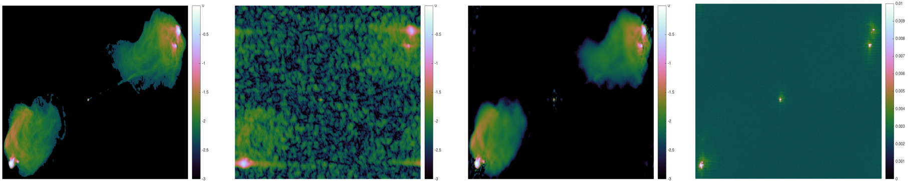


(a) Ground truth (b) Dirty image (c) Mean recovered image

Figure: Cygnus A

Numerical experiments

MYULA with analysis model

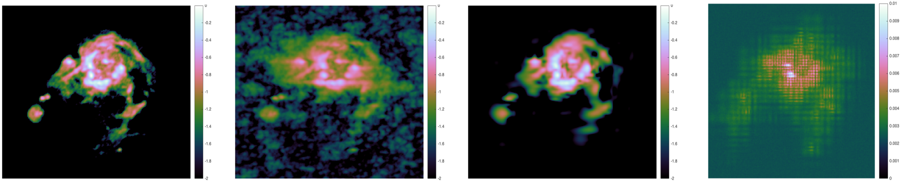


(a) Ground truth (b) Dirty image (c) Mean recovered image (d) Credible interval length

Figure: Cygnus A

Numerical experiments

MYULA with analysis model

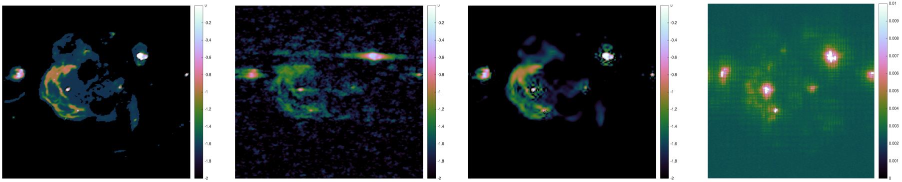


(a) Ground truth (b) Dirty image (c) Mean recovered image (d) Credible interval length

Figure: HII region of M31

Numerical experiments

MYULA with analysis model

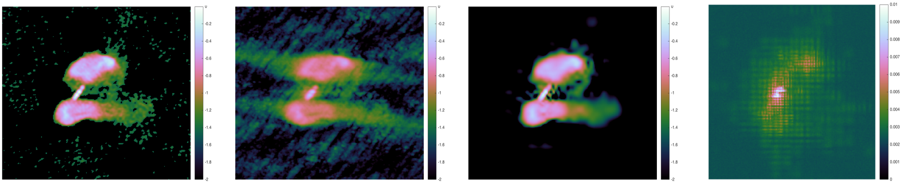


(a) Ground truth (b) Dirty image (c) Mean recovered image (d) Credible interval length

Figure: W28 Supernova remnant

Numerical experiments

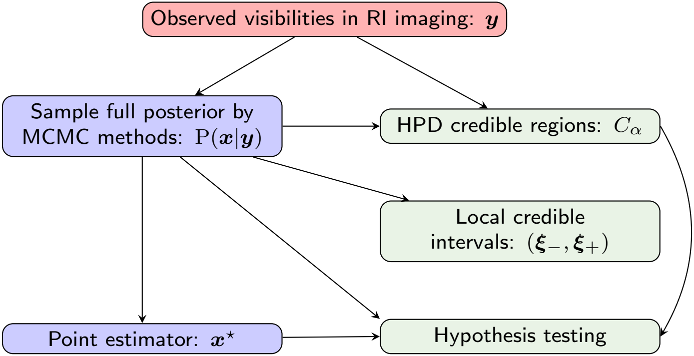
MYULA with analysis model



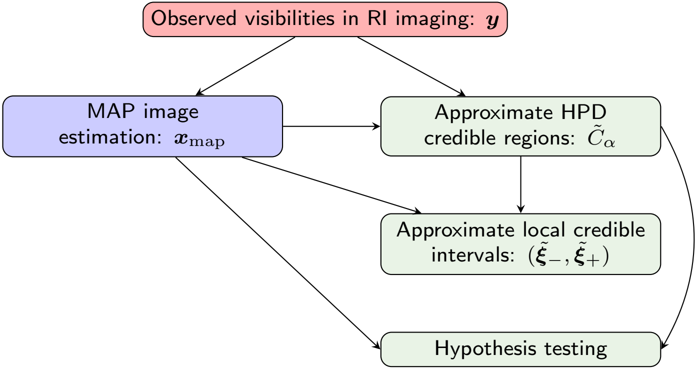
(a) Ground truth (b) Dirty image (c) Mean recovered image (d) Credible interval length

Figure: 3C288

Proximal MCMC sampling and uncertainty quantification



MAP estimation and uncertainty quantification



Approximate Bayesian credible regions for MAP estimation

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

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

where $\tau_\alpha = \sqrt{16\log(3/\alpha)/N}$ and $\alpha \in (4\exp(-N/3), 1)$ (Pereyra 2016b). Follows by recent results from information theory, related to a concentration property of log-concave random vectors.

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

Approximate Bayesian credible regions for MAP estimation

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

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

where $\tau_\alpha = \sqrt{16\log(3/\alpha)/N}$ and $\alpha \in (4\exp(-N/3), 1)$ (Pereyra 2016b). Follows by recent results from information theory, related to a concentration property of log-concave random vectors.

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

Approximate Bayesian credible regions for MAP estimation

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

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

  where $\tau_\alpha = \sqrt{16 \log(3/\alpha)/N}$ and $\alpha \in (4\exp(-N/3), 1)$ (Pereyra 2016b). Follows by recent results from information theory, related to a concentration property of log-concave random vectors.

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

Approximate Bayesian credible regions for MAP estimation

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

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

where $\tau_\alpha = \sqrt{16\log(3/\alpha)/N}$ and $\alpha \in (4\exp(-N/3), 1)$ (Pereyra 2016b). Follows by recent results from information theory, related to a concentration property of log-concave random vectors.

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

Local Bayesian credible intervals for MAP estimation

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).

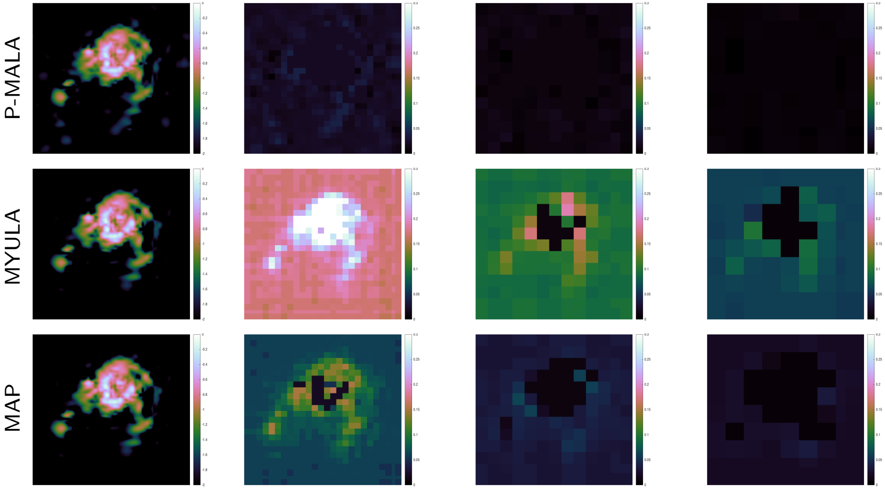
Given $\tilde{\gamma}_\alpha$ and $\boldsymbol{x}^\star$, compute the credible interval by

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

where

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

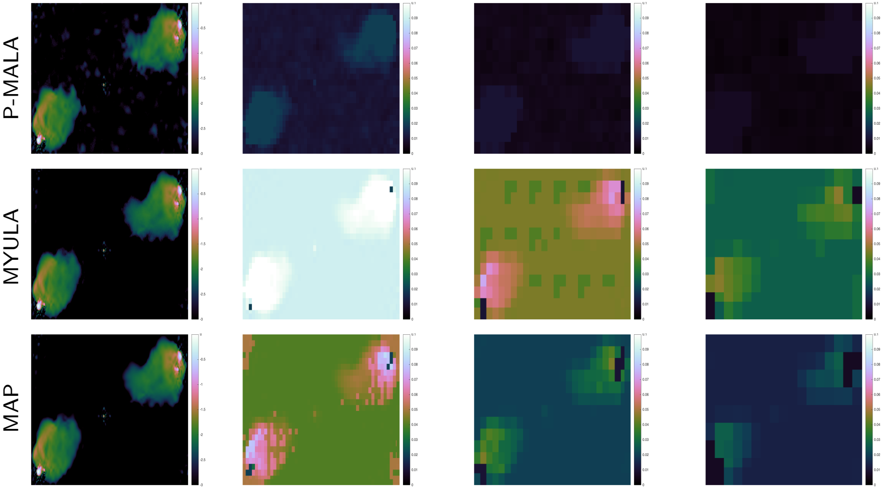
Numerical experiments



(a) point estimators (b) local credible interval grid size 10×10 pixels (c) local credible interval grid size 20×20 pixels (d) local credible interval grid size 30×30 pixels

Figure: Local credible interval computation for M31 for the analysis model.

Numerical experiments



(a) point estimators

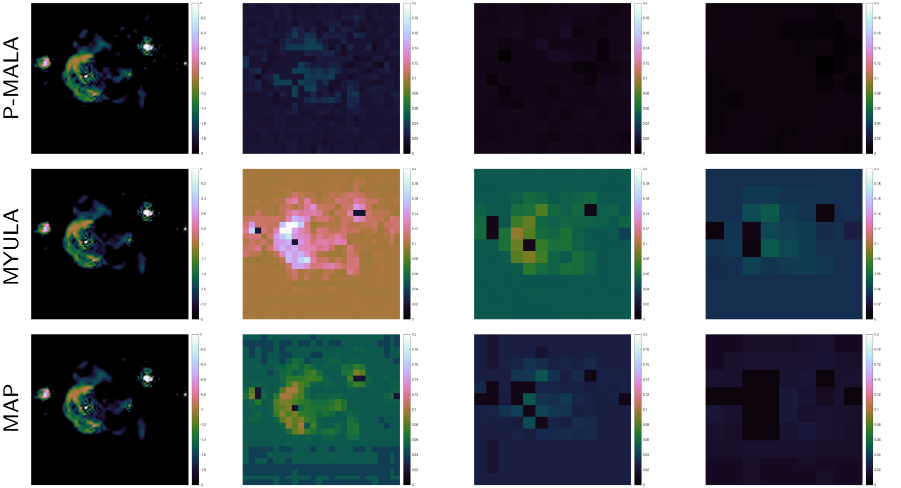
(b) local credible interval grid size 10×10 pixels

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

(d) local credible interval grid size 30×30 pixels

Figure: Local credible interval computation for Cygnus A for the analysis model.

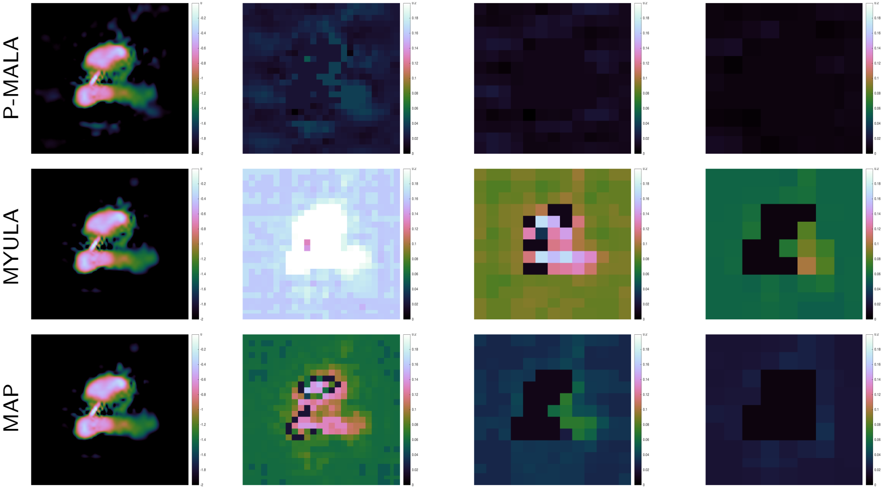
Numerical experiments



(a) point estimators (b) local credible interval grid size 10×10 pixels (c) local credible interval grid size 20×20 pixels (d) local credible interval grid size 30×30 pixels

Figure: Local credible interval computation for W28 for the analysis model.

Numerical experiments



(a) point estimators	(b) local credible interval grid size 10×10 pixels	(c) local credible interval grid size 20×20 pixels	(d) local credible interval grid size 30×30 pixels

Figure: Local credible interval computation for 3C288 for the analysis model.

Numerical experiments

Computation time

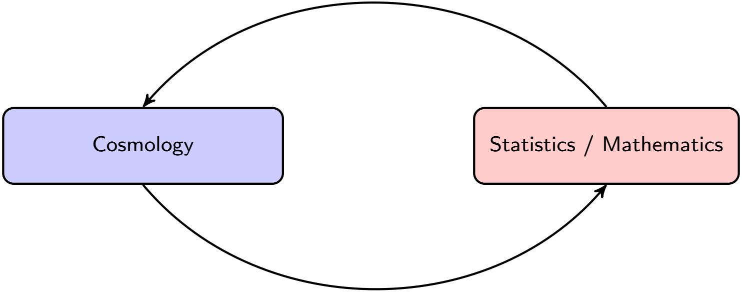
Table: CPU time in minutes for Proximal MCMC sampling and MAP estimation

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

Summary

Closing the DIS loop

Extracting weak observational signatures of fundamental physics from complex data-sets requires sensitive, robust and principled analysis techniques.



Constructing appropriate analysis techniques requires a deep understanding of cosmological problems and methodological foundations.

Extra Slides

Wavelets on the sphere | Wavelets on the ball

E/B separation | Cosmic strings

Analysis vs synthesis | Bayesian interpretations

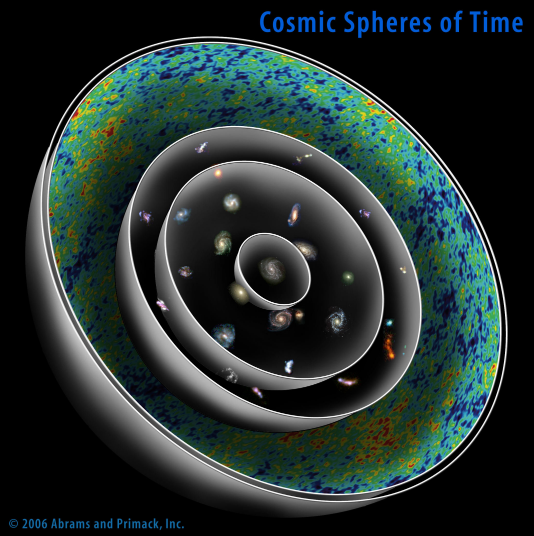
Distribution and parallelisation | PURIFY reconstructions

Proximal MCMC | Hypothesis testing

Extra Slides

Wavelets on the sphere

Observations made on the celestial sphere



Wavelets on the sphere

How can we construct sparsifying transforms?

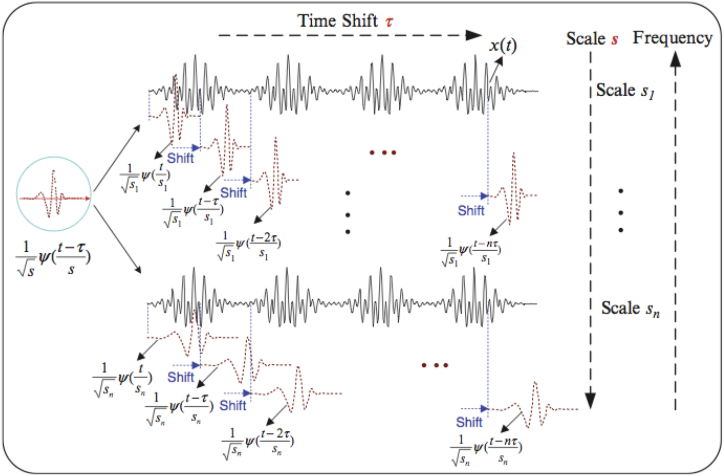


Figure: Wavelet scaling and shifting **[Credit: Gao & Yan (2010)]**

Wavelets on the sphere

Dilation and translation

- Construct wavelet atoms from affine transformations (dilation, translation) on the sphere of a mother wavelet.

- The natural extension of translations to the sphere are rotations. Rotation of a function f on the sphere is defined by

$$\boxed{[\mathcal{R}(\rho)f](\omega) = f(\mathsf{R}_\rho^{-1}\omega), \quad \omega = (\theta, \varphi) \in \mathbb{S}^2, \quad \rho = (\alpha, \beta, \gamma) \in \mathrm{SO}(3) \,.}$$

- How define dilation on the sphere?
  - Stereographic projection
    Antoine & Vandergheynst (1999), Wiaux *et al.* (2005)
  - Harmonic dilation wavelets
    McEwen *et al.* (2006), Sanz *et al.* (2006)
  - Isotropic undecimated wavelets
    Starck *et al.* (2005), Starck *et al.* (2009)
  - Needlets
    Narcowich *et al.* (2006), Baldi *et al.* (2009), Marinucci *et al.* (2008), Geller *et al.* (2008)
  - Scale-discretised wavelets
    Wiaux, McEwen *et al.* (2008), McEwen *et al.* (2003), McEwen *et al.* (2015)

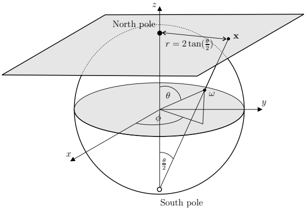


Figure: Stereographic projection

Wavelets on the sphere

Spin scale-discretised wavelet construction

- Spin scale-discretised wavelet ${}_s\Psi^j$ constructed in separable form in harmonic space:

$$\boxed{{}_s\Psi^j_{\ell m} = \kappa^j(\ell)\,\zeta_{\ell m}\ .}$$

- Admissible wavelets constructed to satisfy a resolution of the identity:

$$\underbrace{\boxed{|{}_s\Phi_{\ell 0}|^2}}_{\text{scaling function}} + \sum_{j=0}^{J}\sum_{m=-\ell}^{\ell} \underbrace{\boxed{|{}_s\Psi^j_{\ell m}|^2}}_{\text{wavelet}} = 1\ , \quad \forall \ell\ .$$

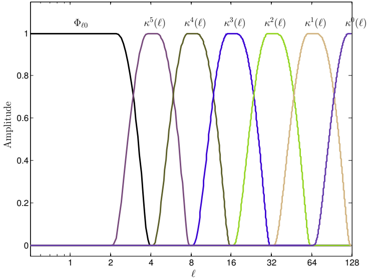
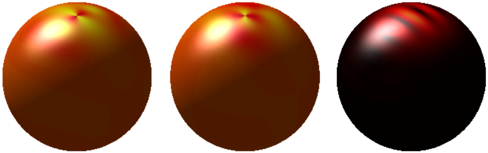


Figure: Harmonic tiling on the sphere.



(a) Real(${}_s\Psi^j$) (b) Imag(${}_s\Psi^j$) (c) Abs(${}_s\Psi^j$)

Figure: Spin scale-discretised wavelets on the sphere.

Wavelets on the sphere

Fast algorithms, variations, and applications

- Fast algorithms critical to scale to large observational data-sets (McEwen *et al.* 2015; McEwen *et al.* 2013; Leistedt, McEwen *et al.* 2013; McEwen *et al.* 2007).

- Variety of types:
  - Spin (McEwen *et al.* 2015)
  - Directional (McEwen *et al.* 2015; Wiaux, McEwen *et al.* 2008)
  - Curvelets (Chan, Leistedt, Kitching & McEwen 2016)
  - Ridgelets (McEwen 2016)
  - Steerable (McEwen *et al.* 2015; Wiaux, McEwen *et al.* 2008)
  - Morphological components (McEwen *et al.* 2008)

- Wavelets ideally suited to cosmological analysis:
  - Physical processes are often manifest on particular physical scales but spatially localised.
  - Localised covariance structure of both theory and data.
  - Observations typically cannot be made over entire celestial sphere.
  - Prevalent CMB analysis technique.

Wavelets on the sphere

Fast algorithms, variations, and applications

- Fast algorithms critical to scale to large observational data-sets (McEwen *et al.* 2015; McEwen *et al.* 2013; Leistedt, McEwen *et al.* 2013; McEwen *et al.* 2007).
- Variety of types:
  - Spin (McEwen *et al.* 2015)
  - Directional (McEwen *et al.* 2015; Wiaux, McEwen *et al.* 2008)
  - Curvelets (Chan, Leistedt, Kitching & McEwen 2016)
  - Ridgelets (McEwen 2016)
  - Steerable (McEwen *et al.* 2015; Wiaux, McEwen *et al.* 2008)
  - Morphological components (McEwen *et al.* 2008)

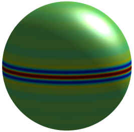


Figure: Ridgelet

- Wavelets ideally suited to cosmological analysis:
  - Physical processes are often manifest on particular physical scales but spatially localised.
  - Localised covariance structure of both theory and data.
  - Observations typically cannot be made over entire celestial sphere.
  - Prevalent CMB analysis technique.

Wavelets on the sphere

Fast algorithms, variations, and applications

- Fast algorithms critical to scale to large observational data-sets (McEwen *et al.* 2015; McEwen *et al.* 2013; Leistedt, McEwen *et al.* 2013; McEwen *et al.* 2007).
- Variety of types:
  - Spin (McEwen *et al.* 2015)
  - Directional (McEwen *et al.* 2015; Wiaux, McEwen *et al.* 2008)
  - Curvelets (Chan, Leistedt, Kitching & McEwen 2016)
  - Ridgelets (McEwen 2016)
  - Steerable (McEwen *et al.* 2015; Wiaux, McEwen *et al.* 2008)
  - Morphological components (McEwen *et al.* 2008)

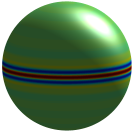


Figure: Ridgelet

- Wavelets ideally suited to cosmological analysis:
  - Physical processes are often manifest on particular physical scales but spatially localised.
  - Localised covariance structure of both theory and data.
  - Observations typically cannot be made over entire celestial sphere.
  - Prevalent CMB analysis technique.

Extra Slides

Wavelets on the ball

Wavelets on the ball (flaglets)

Translation and convolution on the radial line

- Construct translation and convolution on radial line by analogy with infinite line.

- For the standard orthogonal basis $\phi_\omega(x) = \exp^{i\omega x}$ translation of the basis functions defined by shift of coordinates:

$$(\mathcal{T}_u^{\mathbb{R}} \phi_\omega)(x) \equiv \phi_\omega(x-u) = \phi_\omega^*(u)\phi_\omega(x) \ .$$

- Define translation of the spherical Laguerre basis functions on the radial line by analogy:

$$\boxed{(\mathcal{T}_s K_p)(r) \equiv K_p(s) K_p(r) \ .}$$

- Convolution on the radial line defined by

$$\boxed{(f \star h)(r) \equiv \langle f, \mathcal{T}_r h \rangle_{\mathbb{R}+} = \int_{\mathbb{R}+} \mathrm{d}s s^2 \ f(s) \left(\mathcal{T}_r h\right)(s),}$$

- In harmonic space, radial convolution is given by the product

$$(f \star h)_p = \langle f \star h, K_p \rangle_{\mathbb{R}+} = f_p \, h_p \ .$$

Wavelets on the ball (flaglets)

Translation and convolution on the radial line

- Translation on the radial line corresponds to convolution with the Dirac delta:

$$(f \star \delta_s)(r) = \sum_{p=0}^{\infty} f_p K_p(s) K_p(r) = (\mathcal{T}_s f)(r) \; .$$

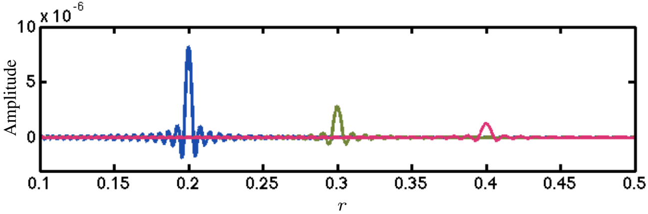


Figure: Band limited translated Dirac delta functions

Wavelets on the ball (flaglets)

Fourier-Laguerre translation and convolution

- Translation operator on the ball defined by combining the angular and radial translation operators, giving

$$\boxed{\mathcal{T}_{\boldsymbol{r}} \equiv \mathcal{T}_r \mathcal{R}_{(\theta,\varphi)} .}$$

- Convolution on the ball of $f \in \mathrm{L}^2(\mathbb{B}^3)$ with an axisymmetric kernel $h \in \mathrm{L}^2(\mathbb{B}^3)$ is defined by

$$\boxed{(f \star h)(\boldsymbol{r}) \equiv \langle f, \mathcal{T}_{\boldsymbol{r}} h \rangle_{\mathbb{B}^3} = \int_{\mathbb{B}^3} \mathrm{d}^3\boldsymbol{s}\; f(\boldsymbol{s}) (\mathcal{T}_{\boldsymbol{r}} h)^*(\boldsymbol{s}),}$$

  where $\boldsymbol{s} \in \mathbb{B}^3$.

- In harmonic space, axisymmetric convolution on the ball may be written

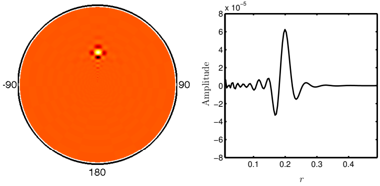
$$(f \star h)_{\ell m p} = \langle f \star h | Z_{\ell m p} \rangle_{\mathbb{B}^3} = \sqrt{\frac{4\pi}{2\ell+1}} f_{\ell m p}\, h^*_{\ell 0 p},$$

  with $f_{\ell m p} = \langle f, Z_{\ell m p} \rangle_{\mathbb{B}^3}$ and $h_{\ell 0 p} \delta_{m0} = \langle h, Z_{\ell m p} \rangle_{\mathbb{B}^3}$.

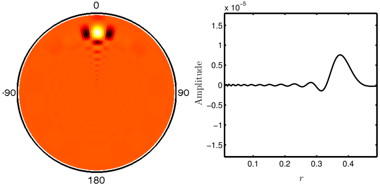
Wavelets on the ball (flaglets)

Fourier-Laguerre translation and convolution

- Angular (radial) aperture of localised functions is invariant under radial (angular) translation.



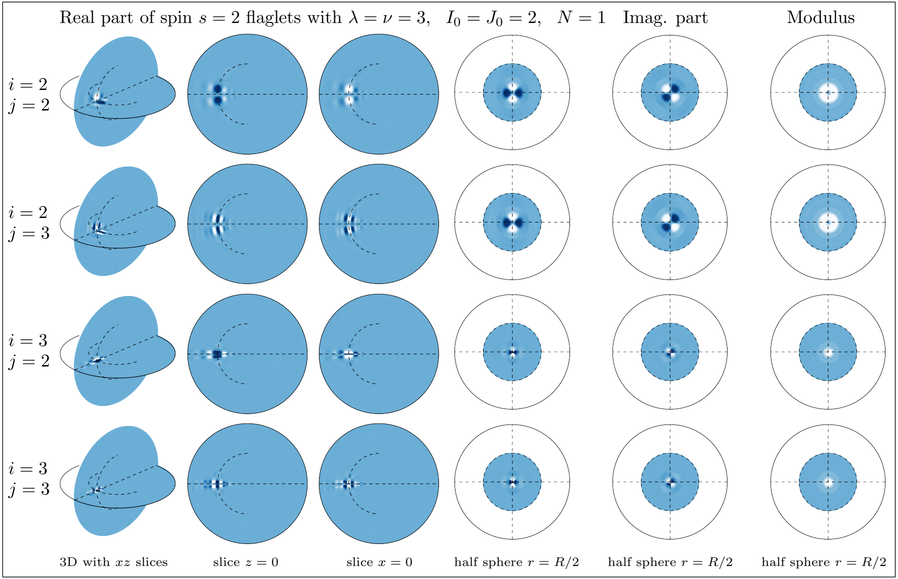
(a) Wavelet kernel translated by $r = 0.2$



(b) Wavelet kernel translated by $r = 0.4$

Figure: Slices of an axisymmetric flaglet wavelet kernel plotted on the ball of radius $R = 0.5$.

Wavelets on the ball (flaglets)



Wavelets on the ball (flaglets)

Wavelet tiling

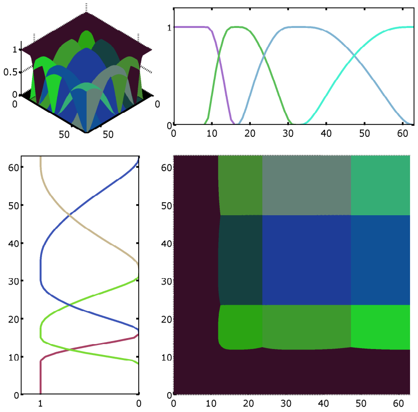


Figure: Tiling of Fourier-Laguerre space.

Extra Slides

E/B separation

E/B separation

Cosmological spin signals

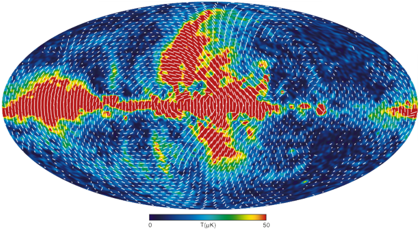
- Observe spin ± 2 cosmological signals on the celestial sphere, with $\boldsymbol{n} = (\theta, \varphi) \in \mathbb{S}^2$:

$$\boxed{{}_{\pm 2}P(\boldsymbol{n}) = Q \pm iU}$$

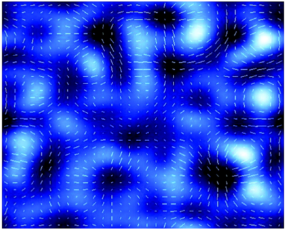
CMB polarization

$$\boxed{{}_{\pm 2}\gamma(\boldsymbol{n}, r) = \gamma_1 \pm i\gamma_2}$$

Cosmic shear



(a) CMB polarization **[Credit: WMAP]**



(b) Cosmic shear **[Credit: Ellis (2010)]**

Figure: Cosmological spin signals.

E/B separation

Parity even and odd components

- Decompose ${}_{\pm 2}P$ into parity even and parity odd components:

$$\boxed{\epsilon(\boldsymbol{n}) = -\frac{1}{2}\left[\bar{\eth}^2 \ {}_2P(\boldsymbol{n}) + \eth^2 \ {}_{-2}P(\boldsymbol{n})\right]}$$

E-mode

$$\boxed{\beta(\boldsymbol{n}) = \frac{\mathrm{i}}{2}\left[\bar{\eth}^2 \ {}_2P(\boldsymbol{n}) - \eth^2 \ {}_{-2}P(\boldsymbol{n})\right]}$$

B-mode

where $\bar{\eth}$ and $\eth$ are spin lowering and raising (differential) operators, respectively.

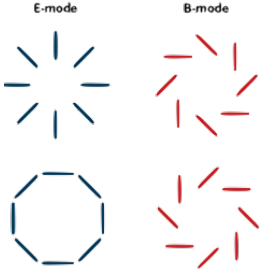


Figure: E-mode (even parity) and B-mode (odd parity) signals **[Credit:** `http://www.skyandtelescope.com/`**]**.

- Different physical processes exhibit different symmetries.

- Can exploit this property to separate signals arising from different underlying physical mechanisms.

E/B separation

Parity even and odd components

- Decompose ${}_{\pm 2}P$ into parity even and parity odd components:

$$\boxed{\epsilon(\boldsymbol{n}) = -\frac{1}{2}\left[\bar{\eth}^2 \; {}_{2}P(\boldsymbol{n}) + \eth^2 \; {}_{-2}P(\boldsymbol{n})\right]} \quad \text{E-mode}$$

$$\boxed{\beta(\boldsymbol{n}) = \frac{\mathrm{i}}{2}\left[\bar{\eth}^2 \; {}_{2}P(\boldsymbol{n}) - \eth^2 \; {}_{-2}P(\boldsymbol{n})\right]} \quad \text{B-mode}$$

where $\bar{\eth}$ and $\eth$ are spin lowering and raising (differential) operators, respectively.

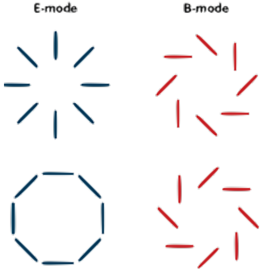


Figure: E-mode (even parity) and B-mode (odd parity) signals **[Credit**: http://www.skyandtelescope.com/**]**.

- Different physical processes exhibit different symmetries.

- Can exploit this property to separate signals arising from different underlying physical mechanisms.

E/B separation

Pure mode wavelet estimator

- On a manifold with boundary (*i.e.* partial sky), E/B decomposition not unique.

- Pure mode wavelet estimators (Leistedt, McEwen, Büttner & Peiris 2016):

$$\widehat{W}_{\epsilon}^{0\Psi^j}(\rho) = -\,\mathrm{Re}\left[\underbrace{\boxed{W^{\pm 2\Upsilon^j}_{\pm 2\widetilde{P}}(\rho)}}_{\text{pseudo}} + \underbrace{\boxed{2W^{\pm 1\Upsilon^j}_{\pm 1\widetilde{P}}(\rho) + W^{0\Upsilon^j}_{0\widetilde{P}}(\rho)}}_{\text{pure correction}}\right],$$

$$\widehat{W}_{\beta}^{0\Psi^j}(\rho) = \mp\,\mathrm{Im}\left[\underbrace{\boxed{W^{\pm 2\Upsilon^j}_{\pm 2\widetilde{P}}(\rho)}}_{\text{pseudo}} + \underbrace{\boxed{2W^{\pm 1\Upsilon^j}_{\pm 1\widetilde{P}}(\rho) + W^{0\Upsilon^j}_{0\widetilde{P}}(\rho)}}_{\text{pure correction}}\right].$$

- Correction terms require spin ± 1 wavelet transforms (McEwen *et al.* 2015).

E/B separation

Pure mode wavelet estimator

- On a manifold with boundary (*i.e.* partial sky), E/B decomposition not unique.

- Pure mode wavelet estimators (Leistedt, McEwen, Büttner & Peiris 2016):

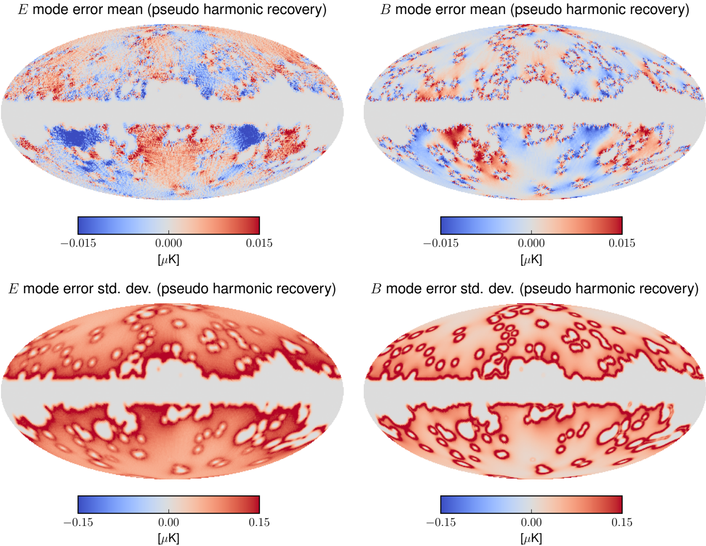
$$\widehat{W}_{\epsilon}^{0\Psi^j}(\rho) = -\,\mathrm{Re}\left[\underbrace{\boxed{W^{\pm 2}_{\pm 2}{}^{\Upsilon^j}_{\widetilde{P}}(\rho)}}_{\text{pseudo}} + \underbrace{\boxed{2W^{\pm 1}_{\pm 1}{}^{\Upsilon^j}_{\widetilde{P}}(\rho) + W^{0}_{0}{}^{\Upsilon^j}_{\widetilde{P}}(\rho)}}_{\text{pure correction}}\right],$$

$$\widehat{W}_{\beta}^{0\Psi^j}(\rho) = \mp\,\mathrm{Im}\left[\underbrace{\boxed{W^{\pm 2}_{\pm 2}{}^{\Upsilon^j}_{\widetilde{P}}(\rho)}}_{\text{pseudo}} + \underbrace{\boxed{2W^{\pm 1}_{\pm 1}{}^{\Upsilon^j}_{\widetilde{P}}(\rho) + W^{0}_{0}{}^{\Upsilon^j}_{\widetilde{P}}(\rho)}}_{\text{pure correction}}\right].$$

- Correction terms require spin ± 1 wavelet transforms (McEwen *et al.* 2015).

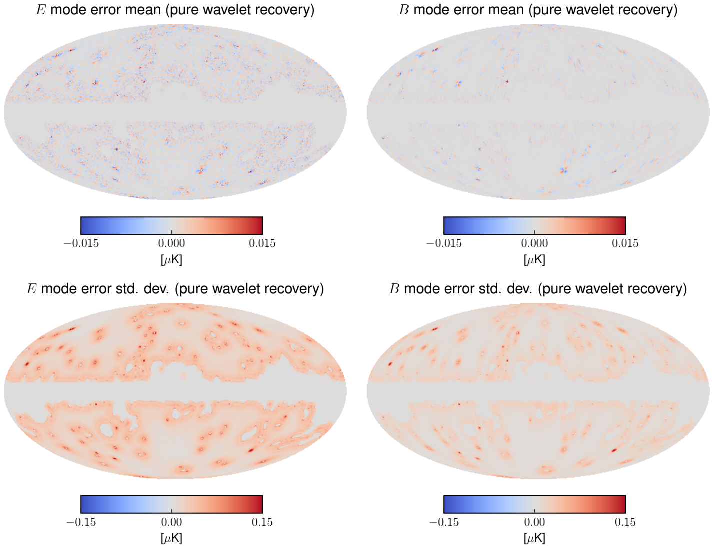
E/B separation

Results: pseudo harmonic approach



E/B separation

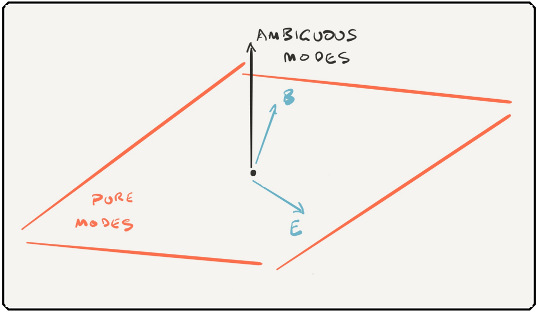
Results: pure wavelet approach



E/B separation

Pure and ambiguous modes

- Pure and ambiguous modes
  (Lewis *et al.* 2002, Bunn *et al.* 2003, Smith 2006, Smith & Zaldarriaga 2007, Grain *et al.* 2007, Ferté *et al.* 2013)
  - E-modes: vanishing curl
  - B-modes: vanishing divergence
  - Pure E-modes: orthogonal to all B-modes
  - Pure B-modes: orthogonal to all E-modes



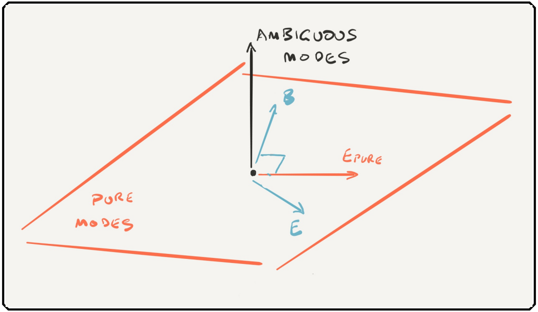
E/B separation

Pure and ambiguous modes

- Pure and ambiguous modes

  (Lewis *et al.* 2002, Bunn *et al.* 2003, Smith 2006, Smith & Zaldarriaga 2007, Grain *et al.* 2007, Ferté *et al.* 2013)

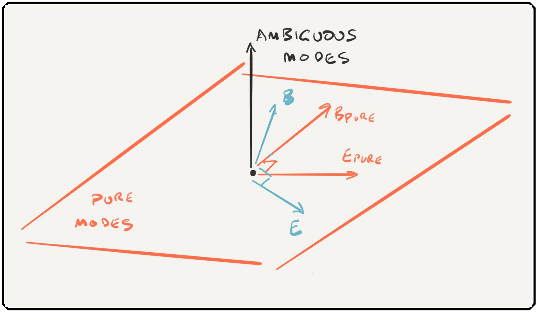
  - E-modes: vanishing curl
  - B-modes: vanishing divergence
  - Pure E-modes: orthogonal to all B-modes
  - Pure B-modes: orthogonal to all E-modes



E/B separation

Pure and ambiguous modes

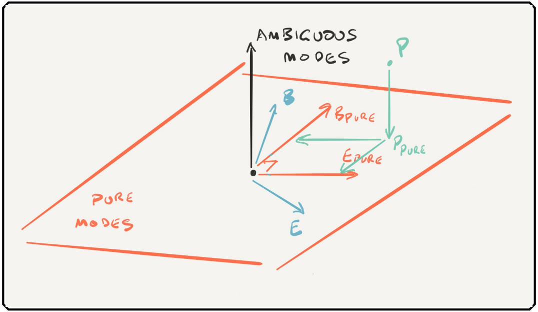
- Pure and ambiguous modes
  (Lewis *et al.* 2002, Bunn *et al.* 2003, Smith 2006, Smith & Zaldarriaga 2007, Grain *et al.* 2007, Ferté *et al.* 2013)
  - E-modes: vanishing curl
  - B-modes: vanishing divergence
  - Pure E-modes: orthogonal to all B-modes
  - Pure B-modes: orthogonal to all E-modes



E/B separation

Pure and ambiguous modes

- Pure and ambiguous modes
  (Lewis *et al.* 2002, Bunn *et al.* 2003, Smith 2006, Smith & Zaldarriaga 2007, Grain *et al.* 2007, Ferté *et al.* 2013)
  - E-modes: vanishing curl
  - B-modes: vanishing divergence
  - Pure E-modes: orthogonal to all B-modes
  - Pure B-modes: orthogonal to all E-modes



E/B separation

Connections between spin and scalar wavelet coefficients

- Spin wavelet transform of ${}_{\pm 2}P = Q \pm \mathrm{i}U$ (observable):

$$\underbrace{\boxed{W^{{}_2\Psi^j}_{{}_{\pm 2}P}(\rho) = \langle {}_{\pm 2}P,\, \mathcal{R}_\rho\, {}_{\pm 2}\Psi^j \rangle}}_{\text{spin wavelet transform}} = \int_{\mathbb{S}^2} \mathrm{d}\Omega(\omega)\, {}_{\pm 2}P(\omega) (\mathcal{R}_\rho\, {}_{\pm 2}\Psi^j)^*(\omega)\,.$$

- Scalar wavelet transforms of E and B (non-observable):

$$\underbrace{\boxed{W^{{}_0\Psi^j}_{\epsilon}(\rho) = \langle \epsilon,\, \mathcal{R}_\rho\, {}_0\Psi^j \rangle}}_{\text{scalar wavelet transform}}\,,$$

$$\underbrace{\boxed{W^{{}_0\Psi^j}_{\beta}(\rho) = \langle \beta,\, \mathcal{R}_\rho\, {}_0\Psi^j \rangle}}_{\text{scalar wavelet transform}}\,,$$

where ${}_0\Psi^j \equiv \bar{\eth}^2\, {}_2\Psi^j$.

- Spin wavelet coefficients of ${}_{\pm 2}P$ are connected to scalar wavelet coefficients of E/B:

$$\boxed{W^{{}_0\Psi^j}_{\epsilon}(\rho) = -\mathrm{Re}\Big[W^{{}_2\Psi^j}_{{}_{\pm 2}P}(\rho)\Big] \quad \text{and} \quad W^{{}_0\Psi^j}_{\beta}(\rho) = \mp\mathrm{Im}\Big[W^{{}_2\Psi^j}_{{}_{\pm 2}P}(\rho)\Big]\,.}$$

E/B separation
Exploiting wavelets

General approach to recover E/B signals using scale-discretised wavelets

1. Compute spin wavelet transform of ${}_{\pm 2}P = Q + \mathrm{i}U$:

$$ {}_{\pm 2}P(\omega) \xrightarrow[\text{S2LET}]{\text{Spin wavelet transform}} W^{{}_2\Psi^j}_{{}_{\pm 2}P}(\rho) $$

2. Account for mask in wavelet domain (simultaneous harmonic and spatial localisation):

$$ W^{{}_2\Psi^j}_{{}_{\pm 2}P}(\rho) \xrightarrow{\text{Mitigate mask}} \bar{W}^{{}_2\Psi^j}_{{}_{\pm 2}P}(\rho) $$

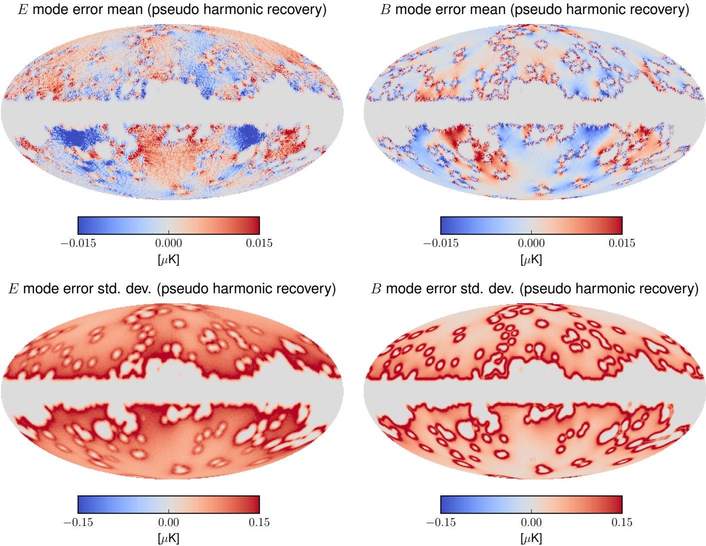
3. Construct E/B maps:

(a) $$ W^{{}_0\Psi^j}_{\epsilon}(\rho) = -\mathrm{Re}\left[\bar{W}^{{}_2\Psi^j}_{{}_{\pm 2}P}(\rho)\right] \xrightarrow[\text{S2LET}]{\text{Inverse scalar wavelet transform}} \epsilon(\omega) $$

(b) $$ W^{{}_0\Psi^j}_{\beta}(\rho) = \mp\mathrm{Im}\left[\bar{W}^{{}_2\Psi^j}_{{}_{\pm 2}P}(\rho)\right] \xrightarrow[\text{S2LET}]{\text{Inverse scalar wavelet transform}} \beta(\omega) $$

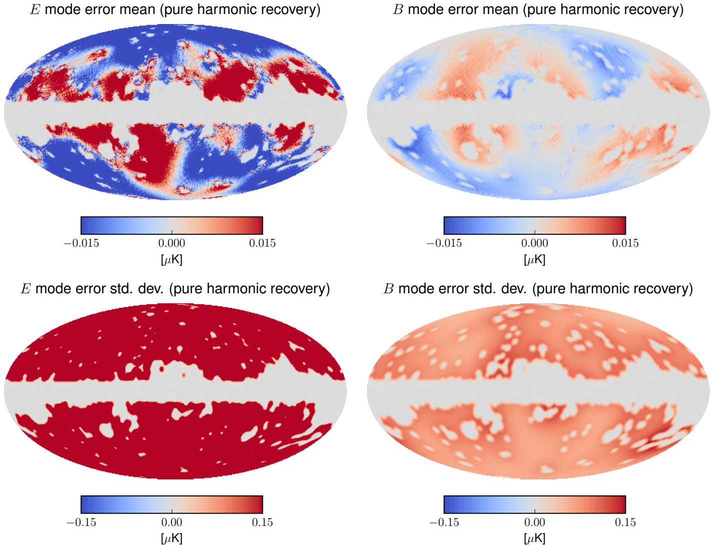
E/B separation

Results: pseudo harmonic approach



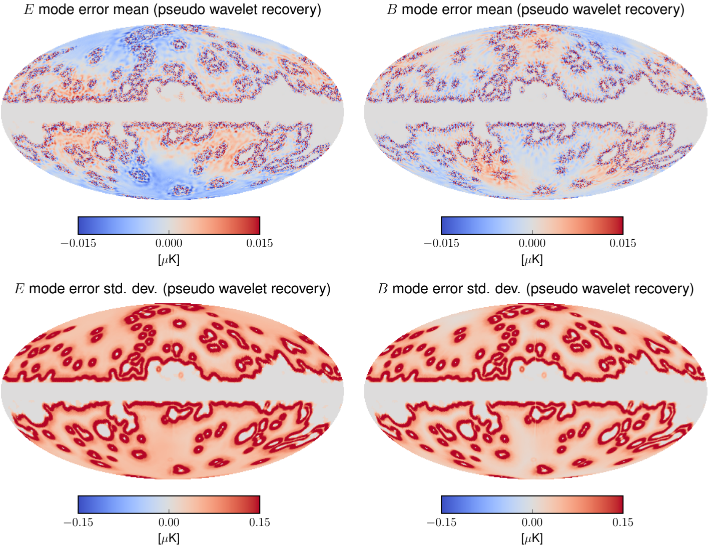
E/B separation

Results: pure harmonic approach



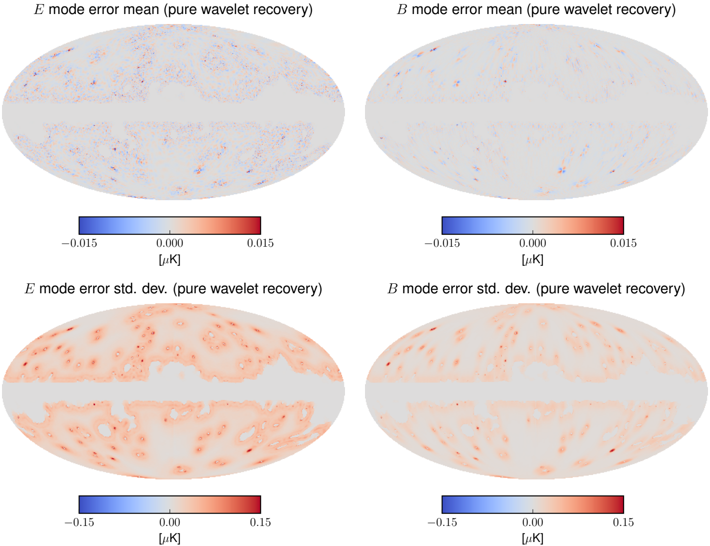
E/B separation

Results: pseudo wavelet approach



E/B separation

Results: pure wavelet approach



Extra Slides

Cosmic strings

Cosmic strings
Wavelet space distributions

- Calculate analytically the probability distribution of the CMB in wavelet space:

$$\mathrm{P}_j^c(W_{j\rho}^c) = \frac{1}{\sqrt{2\pi(\sigma_j^c)^2}} \exp\left(-\frac{1}{2}\left(\frac{W_{j\rho}^c}{\sigma_j^c}\right)^2\right), \text{ where } (\sigma_j^c)^2 = \langle W_{j\rho}^c \, W_{j\rho}^{c\,*}\rangle = \sum_{\ell m} C_\ell \, |(\Psi_j)_{\ell m}|^2 .$$

- Fit a generalised Gaussian distribution (GGD) for the wavelet coefficients of a string training map:

$$\mathrm{P}_j^s(W_{j\rho}^s \,|\, G\mu) = \frac{\upsilon_j}{2G\mu\nu_j\Gamma(\upsilon_j{}^{-1})} \exp\left(-\left|\frac{W_{j\rho}^s}{G\mu\nu_j}\right|^{\upsilon_j}\right),$$

with scale parameter ν_j and shape parameter υ_j.

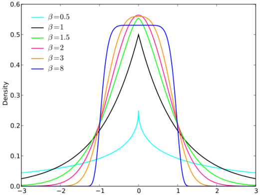
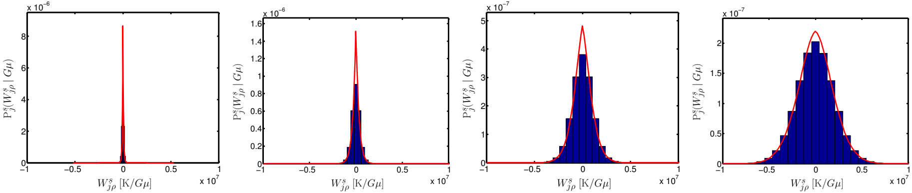


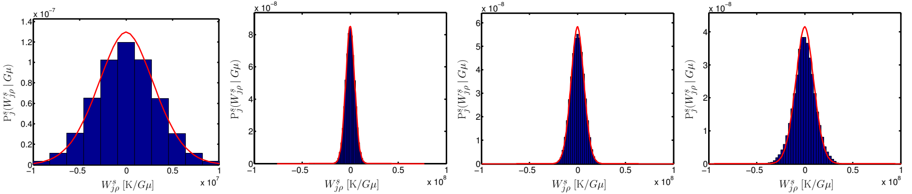
Figure: Generalised Gaussian distribution (GGD).

Cosmic strings

Cosmic string distributions in wavelet space



(a) $j = 0$ (b) $j = 1$ (c) $j = 2$ (d) $j = 3$

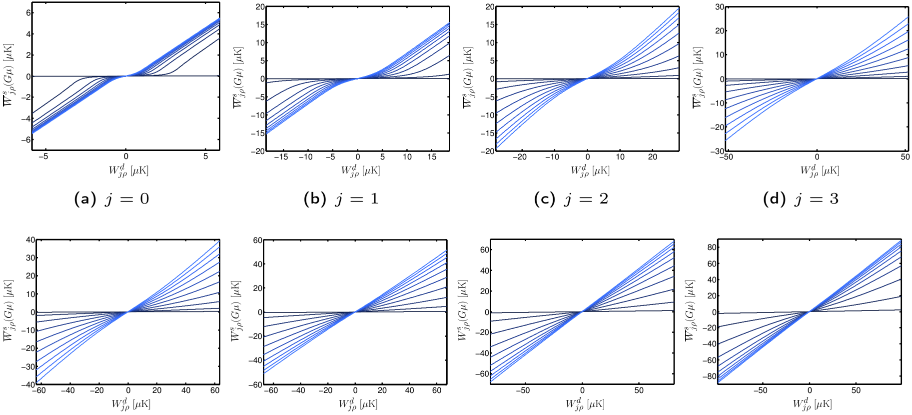


(e) $j = 4$ (f) $j = 5$ (g) $j = 6$ (h) $j = 7$

Figure: Distribution of the cosmic string maps in wavelet space for each wavelet scale j.

Cosmic strings

Cosmic string distributions in wavelet space



(a) $j = 0$ (b) $j = 1$ (c) $j = 2$ (d) $j = 3$

(e) $j = 4$ (f) $j = 5$ (g) $j = 6$ (h) $j = 7$

Figure: Bayesian thresholding functions for each wavelet scale j.

Extra Slides

Analysis vs synthesis

SARA algorithm

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

- Overcomplete dictionary composed of a concatenation of orthonormal bases:

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

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

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

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

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

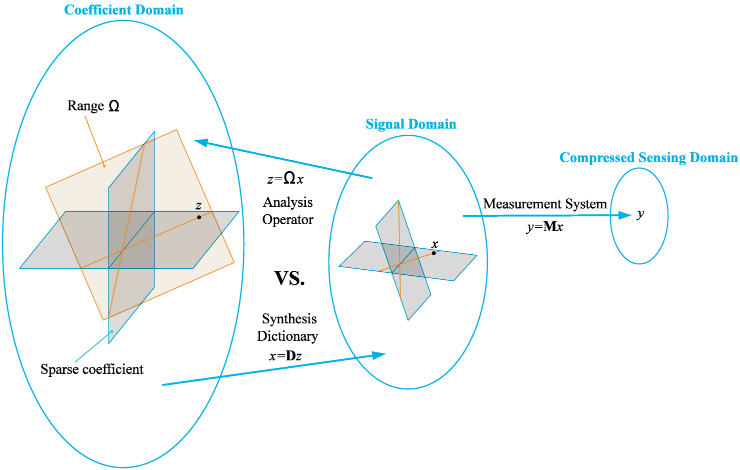


Figure: Analysis- and synthesis-based approaches [Credit: Nam *et al.* (2012)].

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

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

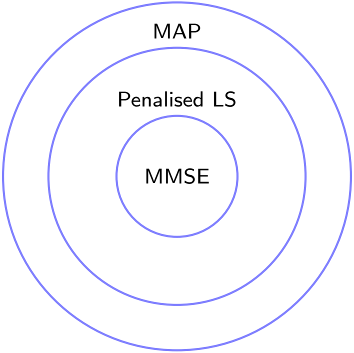
synthesis

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

Bayesian interpretations

Other Bayesian interpretations of the synthesis-based approach

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



Bayesian interpretations

One Bayesian interpretation of the analysis-based approach

- Analysis-based MAP estimate is

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

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

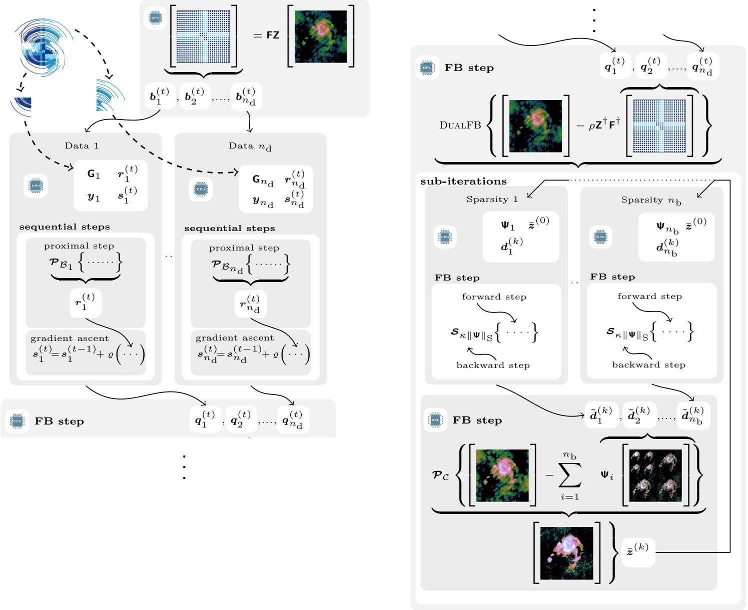
Distributed and parallelised convex optimisation

Distributed and parallelised convex optimisation

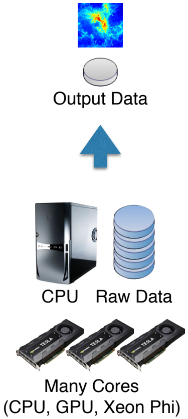
- Solve resulting convex optimisation problems by proximal splitting.
- Block inexact ADMM algorithm to split data and measurement operator:
(Carrillo, McEwen & Wiaux 2014; Onose, Carrillo, Repetti, McEwen, *et al.* 2016)

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

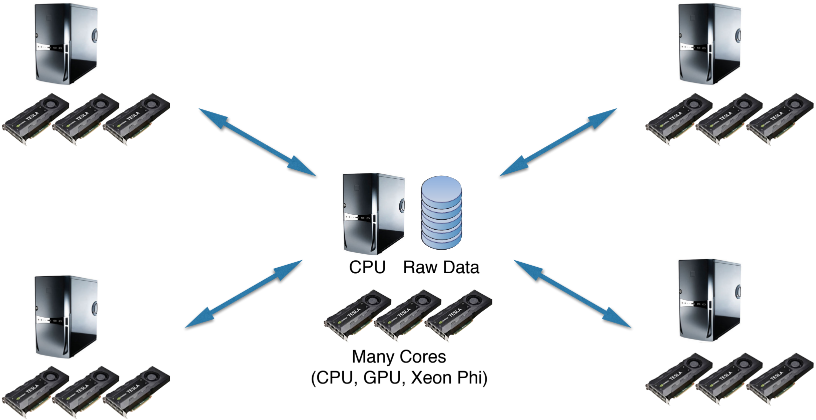
Distributed and parallelised convex optimisation



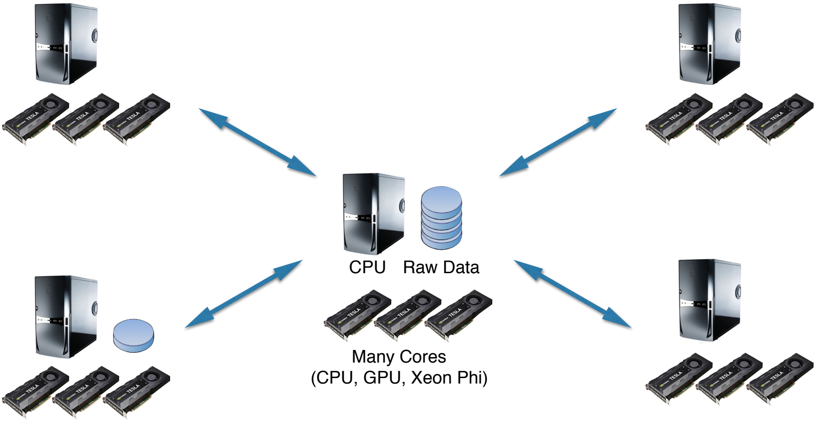
Standard algorithms



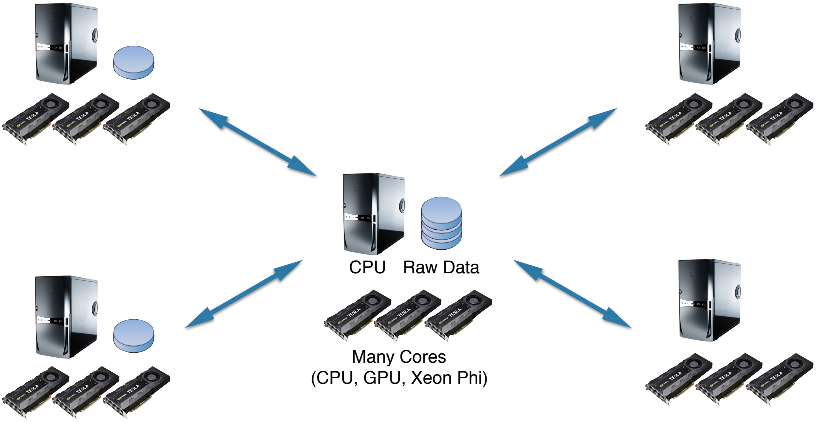
Highly distributed and parallelised algorithms



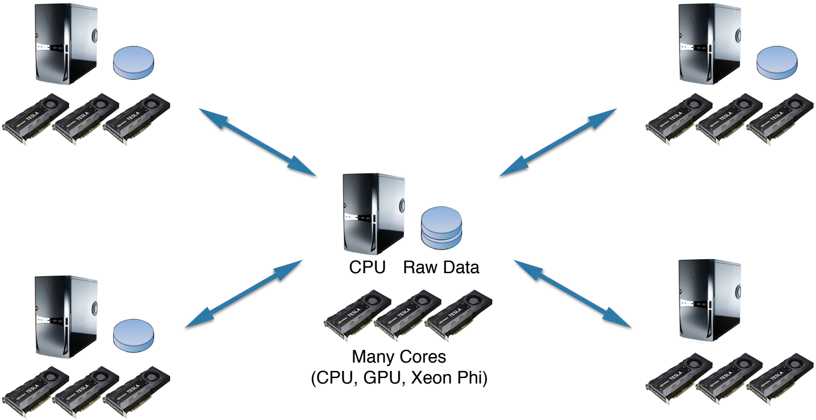
Highly distributed and parallelised algorithms



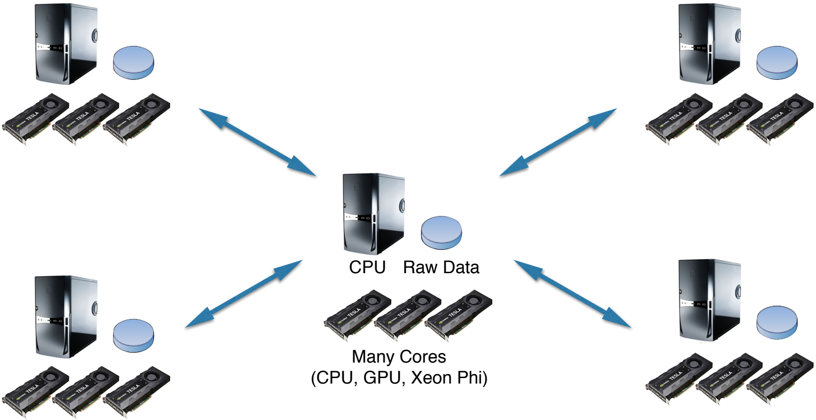
Highly distributed and parallelised algorithms



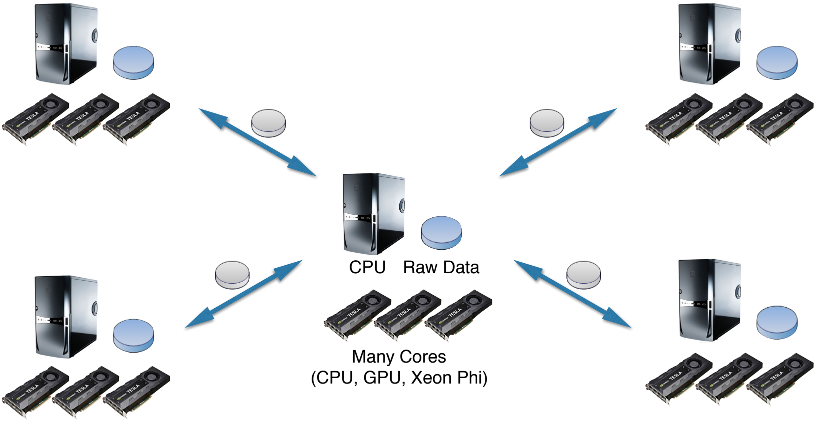
Highly distributed and parallelised algorithms



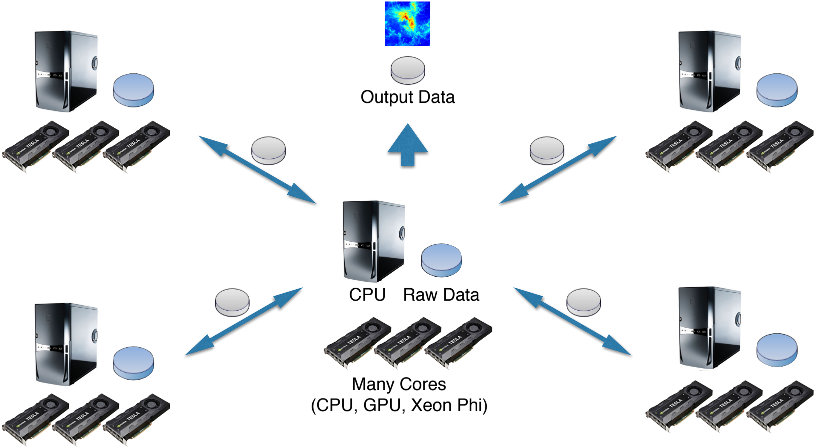
Highly distributed and parallelised algorithms



Highly distributed and parallelised algorithms



Highly distributed and parallelised algorithms



Extra Slides

PURIFY reconstructions

PURIFY reconstruction

VLA observation of 3C129

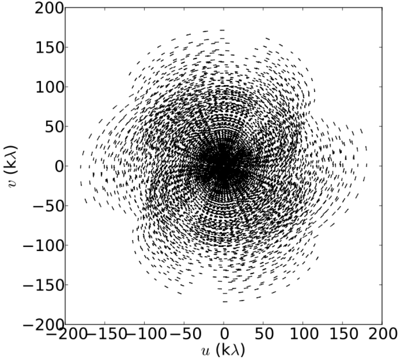
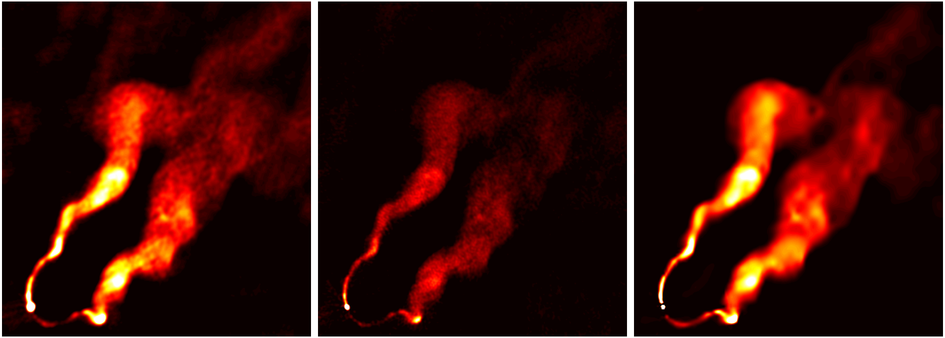


Figure: VLA visibility coverage for 3C129

PURIFY reconstruction

VLA observation of 3C129

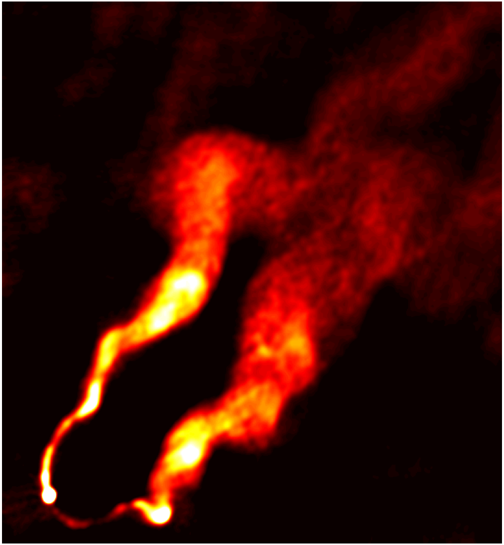


(a) CLEAN (natural) (b) CLEAN (uniform) (c) PURIFY

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

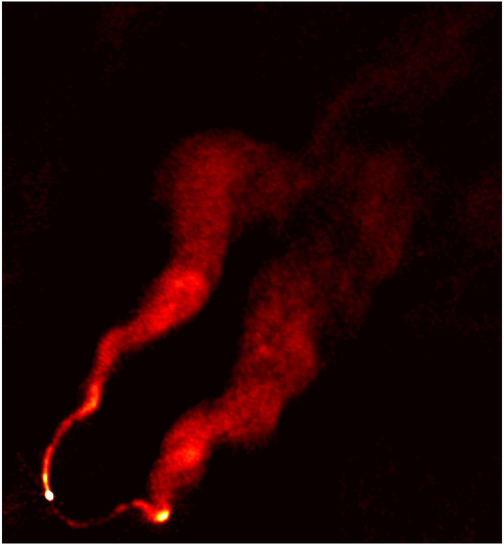
PURIFY reconstruction

VLA observation of 3C129 imaged by CLEAN (natural)



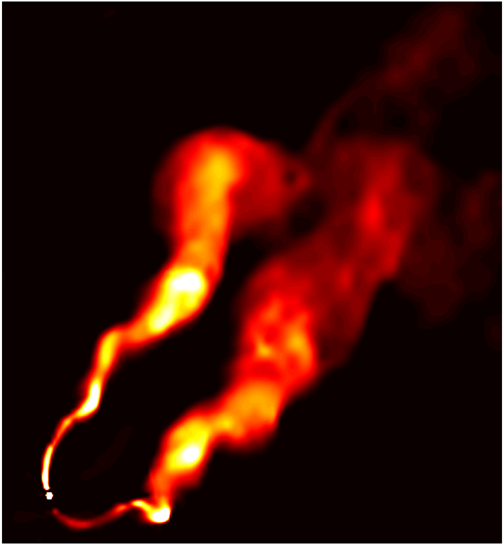
PURIFY reconstruction

VLA observation of 3C129 images by CLEAN (uniform)



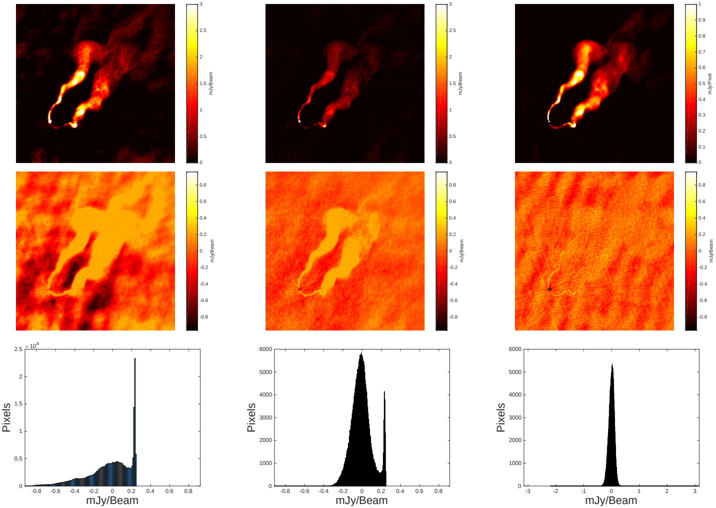
PURIFY reconstruction

VLA observation of 3C129 images by PURIFY



PURIFY reconstruction

VLA observation of 3C129



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

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

PURIFY reconstruction

VLA observation of Cygnus A

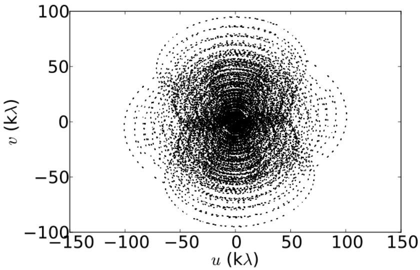
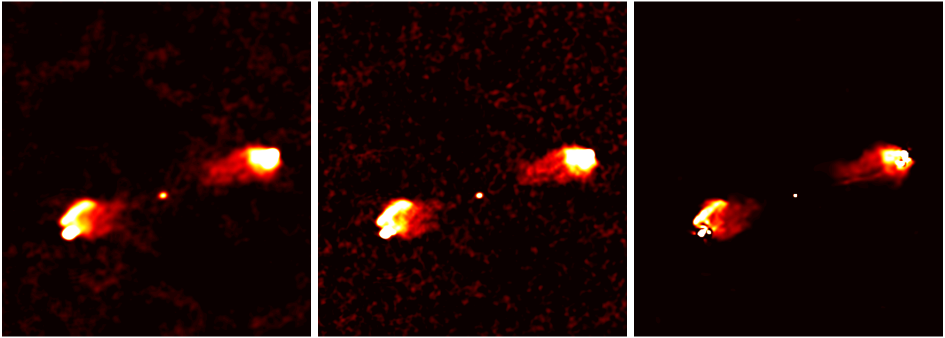


Figure: VLA visibility coverage for Cygnus A

PURIFY reconstruction

VLA observation of Cygnus A

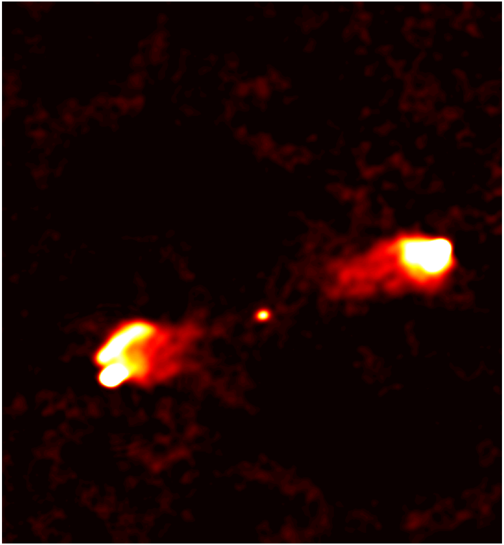


(a) CLEAN (natural) (b) CLEAN (uniform) (c) PURIFY

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

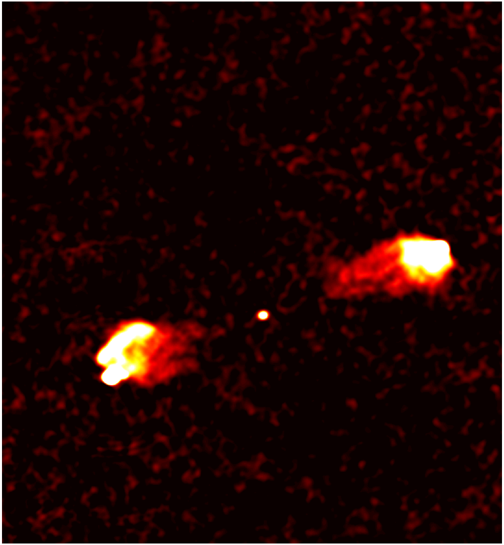
PURIFY reconstruction

VLA observation of Cygnus A imaged by CLEAN (natural)



PURIFY reconstruction

VLA observation of Cygnus A images by CLEAN (uniform)

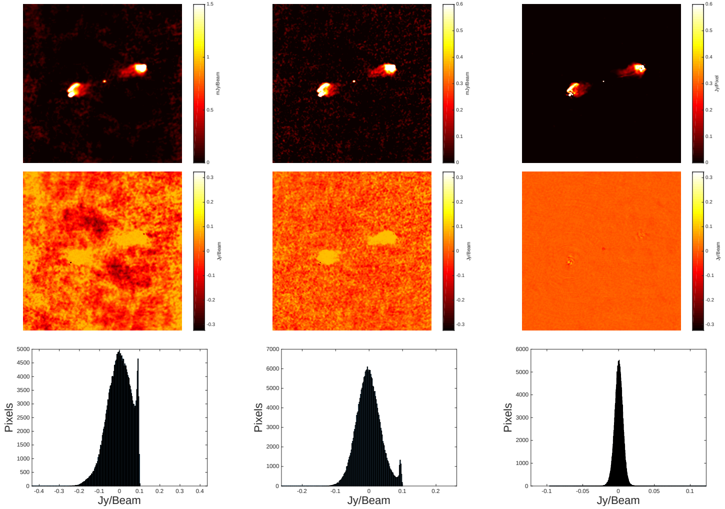


PURIFY reconstruction

VLA observation of Cygnus A images by PURIFY

PURIFY reconstruction

VLA observation of Cygnus A



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

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

PURIFY reconstruction

ATCA observation of PKS J0334-39

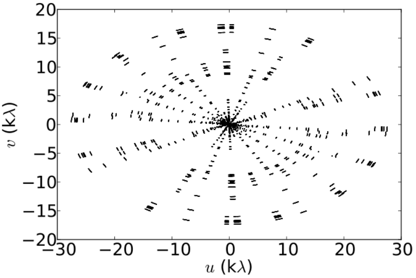
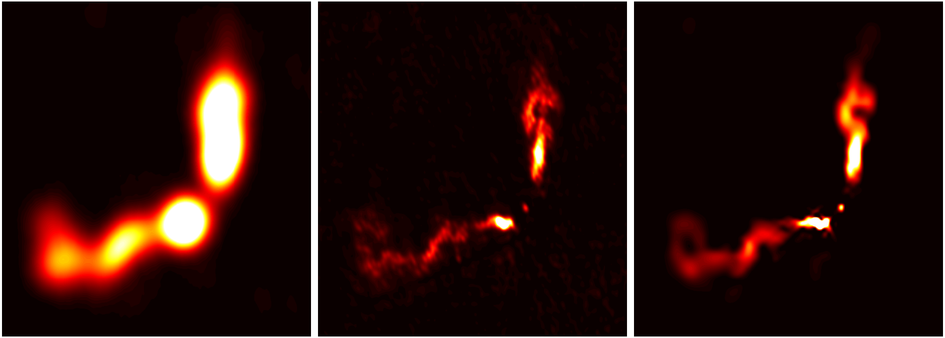


Figure: VLA visibility coverage for PKS J0334-39

PURIFY reconstruction

ATCA observation of PKS J0334-39

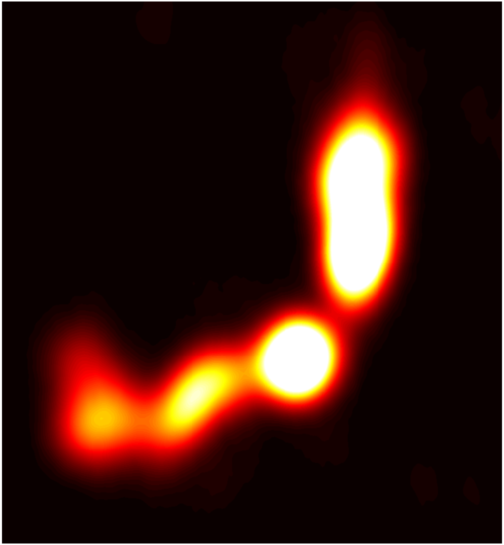


(a) CLEAN (natural) (b) CLEAN (uniform) (c) PURIFY

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

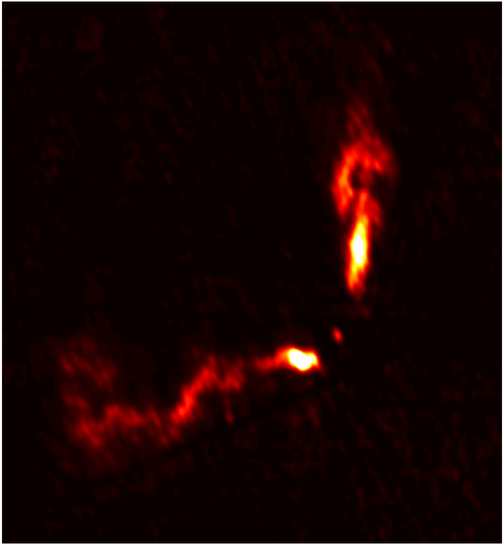
PURIFY reconstruction

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



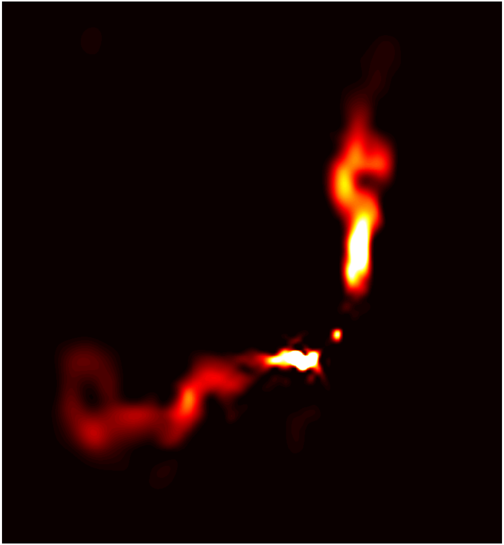
PURIFY reconstruction

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



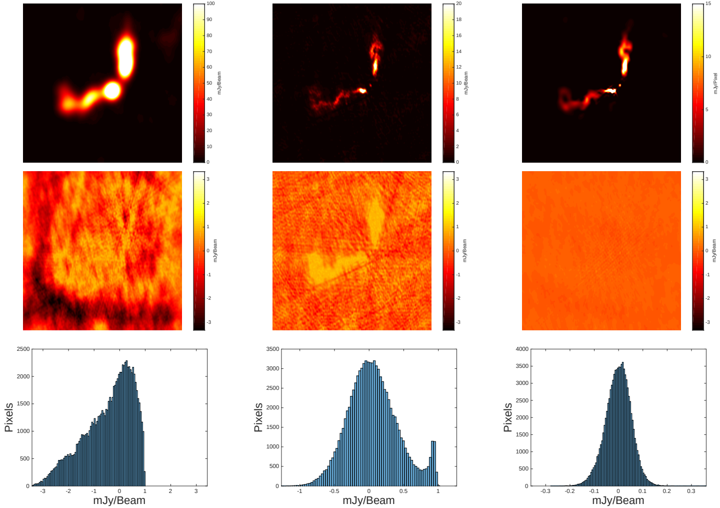
PURIFY reconstruction

VLA observation of PKS J0334-39 images by PURIFY



PURIFY reconstruction

ATCA observation of PKS J0334-39



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

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

PURIFY reconstruction

ATCA observation of PKS J0116-473

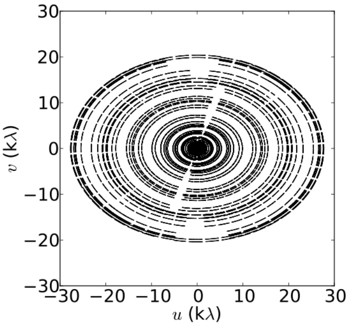
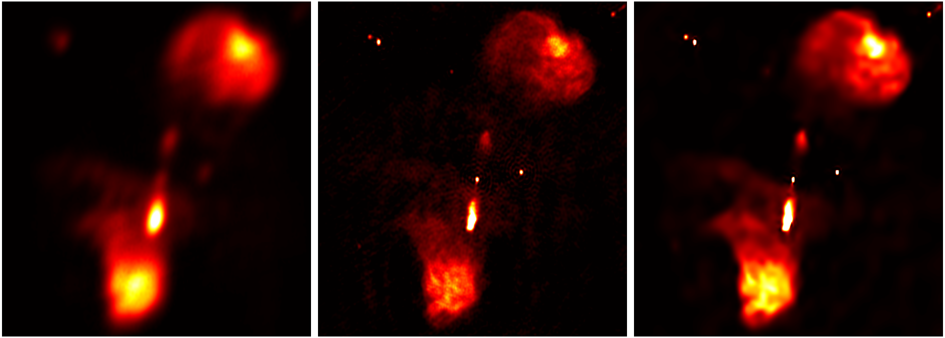


Figure: ATCA visibility coverage for Cygnus A

PURIFY reconstruction

ATCA observation of PKS J0116-473

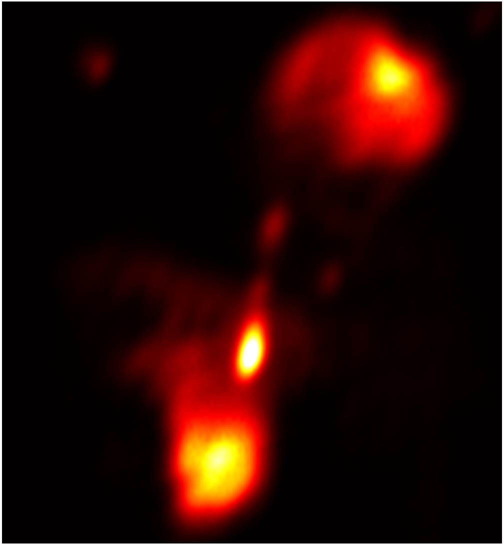


(a) CLEAN (natural) (b) CLEAN (uniform) (c) PURIFY

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

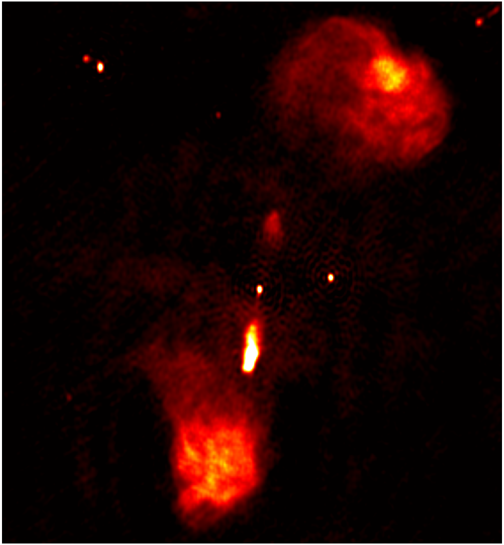
PURIFY reconstruction

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



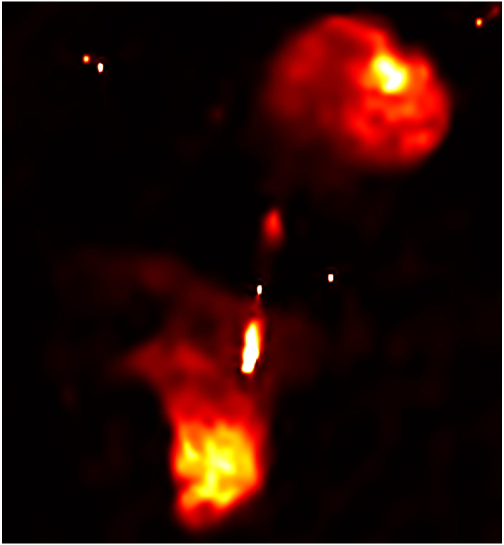
PURIFY reconstruction

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



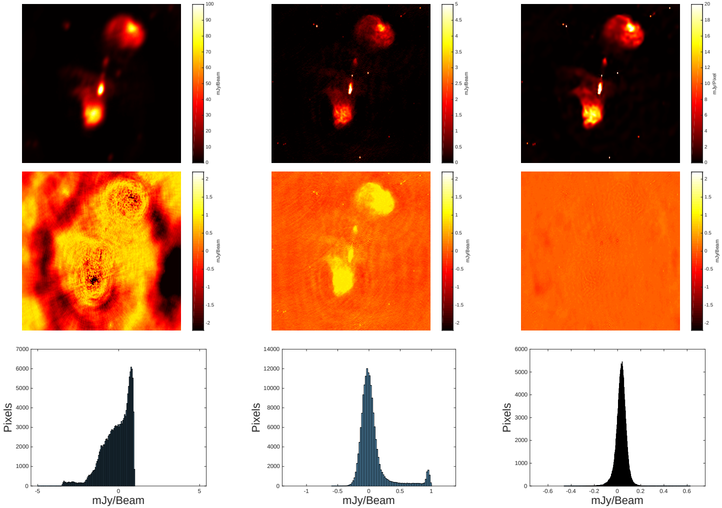
PURIFY reconstruction

VLA observation of PKS J0116-473 images by PURIFY



PURIFY reconstruction

ATCA observation of PKS J0116-473



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

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

PURIFY reconstructions

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

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

Extra Slides

Proximal MCMC

Proximity operators

A brief aside

- Define proximity operator:

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

- Generalisation of projection operator:

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

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

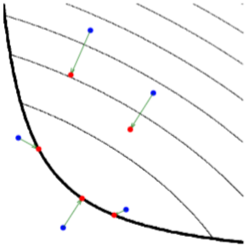


Figure: Illustration of proximity operator [Credit: Parikh & Boyd (2013)]

Proximal MCMC methods

- Exploit proximal calculus.
- "Replace gradients with sub-gradients".

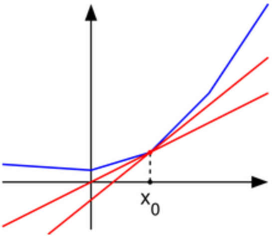


Figure: Illustration of sub-gradients [Credit: Wikipedia (Maksim)]

Proximal MALA

Moreau approximation

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

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

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

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

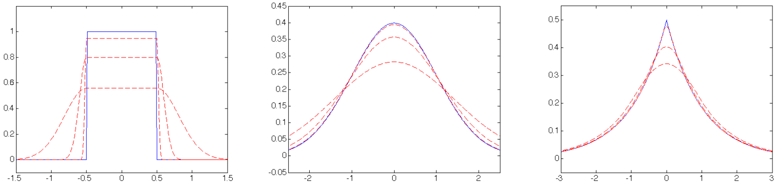


Figure: Illustration of Moreau approximations [Credit: Pereyra 2016a]

Proximal MALA

Computing proximity operators for the analysis case

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

  Prior Likelihood

- Must solve an optimisation problem for each iteration!

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

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

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

Single forward-backward iteration

- Analytic approximation:

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

Proximal MALA

Computing proximity operators for the synthesis case

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

  Prior Likelihood

- Must solve an optimisation problem for each iteration!

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

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

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

Single forward-backward iteration

- Analytic approximation:

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

MYULA

Moreau-Yosida approximation

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

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

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

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

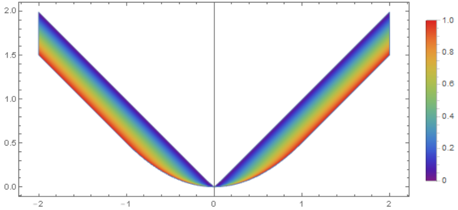


Figure: Illustration of Moreau-Yosida envelope of $|x|$ for varying λ [Credit: Stack exchange (ubpdqn)]

MYULA

Computing proximity operators for the analysis case

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

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

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

MYULA

Computing proximity operators for the synthesis case

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

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

Extra Slides

Hypothesis testing

Hypothesis testing
Method

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

Hypothesis testing of physical structure

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

Hypothesis testing
Method

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

Hypothesis testing of physical structure

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

Hypothesis testing

Method

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

Hypothesis testing of physical structure

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

Hypothesis testing

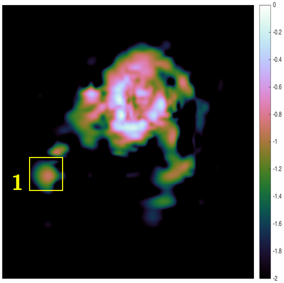
Method

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

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 artifact with confidence $(1-\alpha)\%$, *i.e.* structure most likely physical.
    - If $\boldsymbol{x}' \in C_\alpha$ uncertainly too high to draw strong conclusions about the physical nature of the structure.

Hypothesis testing
Numerical experiments

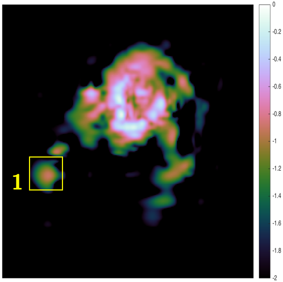


(a) Recovered image

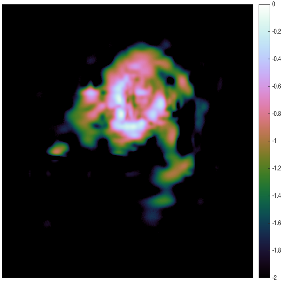
Figure: HII region of M31

Hypothesis testing

Numerical experiments



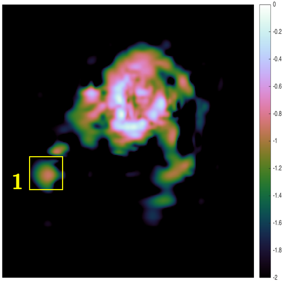
(a) Recovered image



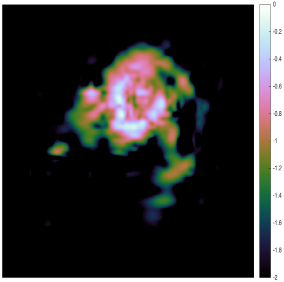
(b) Surrogate with region removed

Figure: HII region of M31

Hypothesis testing
Numerical experiments



(a) Recovered image



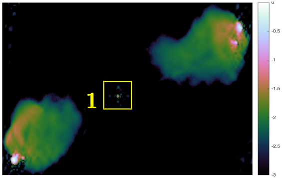
(b) Surrogate with region removed

Figure: HII region of M31

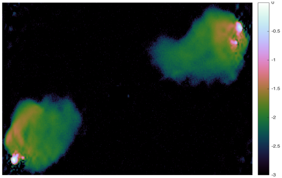
1. Reject null hypothesis

   $\Rightarrow$ structure physical

Hypothesis testing
Numerical experiments



(a) Recovered image



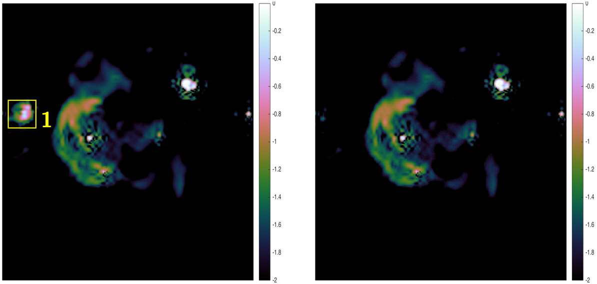
(b) Surrogate with region removed

Figure: Cygnus A

1. Cannot reject null hypothesis

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

Hypothesis testing
Numerical experiments



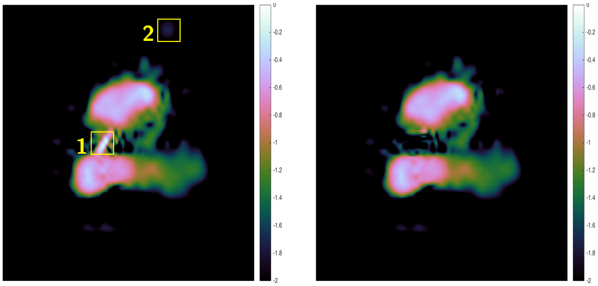
(a) Recovered image (b) Surrogate with region removed

Figure: Supernova remnant W28

1. Reject null hypothesis

$\Rightarrow$ structure physical

Hypothesis testing
Numerical experiments



(a) Recovered image

(b) Surrogate with region removed

Figure: 3C288

1. Reject null hypothesis

   ⇒ structure physical

2. Cannot reject null hypothesis

   ⇒ cannot make strong statistical statement about origin of structure

Hypothesis testing

Comparison of numerical experiments

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

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

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