Scalable imaging algorithms for radioastronomy
- *& story of a new black hole discovery* -

Prof. Yves Wiaux

Institute of Sensors, Signals and Systems,
Heriot-Watt Edinburgh, UK

May 22, 2017

Interferometric imaging in the SKA era

Our previous work highlighted...

- *Convex optimisation* - *"Compressed sensing imaging techniques for radio interferometry"*, MNRAS, 2009
- *Compressive sensing* - *"Spread spectrum for imaging techniques in radio interferometry"*, MNRAS, 2009

- *Software* - *"PURIFY: a new approach to radio-interferometric imaging"*, MNRAS, 2014
- *Google our most recent publications...*

- Increase the resolution and sensitivity up to two orders of magnitude over current instruments

  Terapixel images   huge dynamic range

- Unprecedented amount of data to be processed: orders of magnitude beyond image size.

Measurement model

Inverse problem

- Measurement equation

$$y(\boldsymbol{u}) = \int D(\boldsymbol{l}, \boldsymbol{u}) x(\boldsymbol{l}) e^{-2i\pi \boldsymbol{u}\cdot\boldsymbol{l}} \mathrm{d}^2\boldsymbol{l}$$

- Discretised version of the ill-posed inverse problem

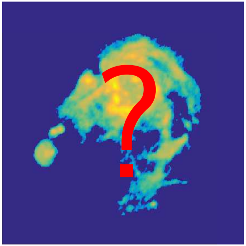
$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{n} \quad \text{with} \quad \boldsymbol{\Phi} = \mathbf{G}\mathbf{F}$$

  - $\boldsymbol{x} \in \mathbb{R}_+^N$ the intensity image of interest
  - $\boldsymbol{\Phi} \in \mathbb{C}^{M \times N}$ a linear map; image domain to visibility space
  - $\boldsymbol{y} \in \mathbb{C}^M$ the measured visibilities
  - $\mathbf{G} \in \mathbb{C}^{M \times kN}$ gridding matrix modelling DDEs
  - $\mathbf{F} \in \mathbb{C}^{kN \times N}$ Fourier matrix with zero padding

Ill-posed inverse problem

SKA *u*–*v* coverage

$= \boldsymbol{\Phi}$



$+ \boldsymbol{n}$

Traditional imaging techniques

CLEAN

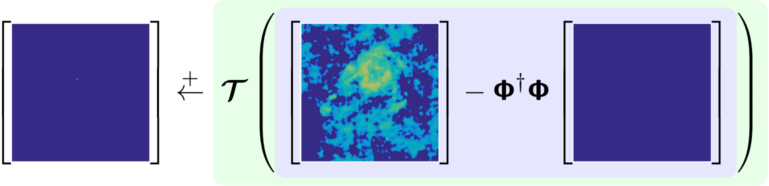
- Greedy iterative deconvolution algorithm
  - Select atoms associated with brightest pixel of residual image
  - Build the solution implicitly imposing sparsity in image space

$$\boldsymbol{x}^{(t)} = \boldsymbol{x}^{(t-1)} + \boldsymbol{\mathcal{T}}\left(\boldsymbol{\Phi}^{\dagger}\left(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}^{(t-1)}\right)\right)$$

- Forward - backward *like* structure
  - Forward step (major cycle) in the gradient direction of the ℓ_2 norm of the residual image
  - Backward step (minor cycle) with non-linear sparsity-enforcing operator $\boldsymbol{\mathcal{T}}$

Traditional imaging techniques

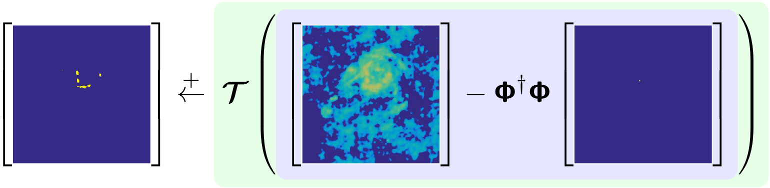
CLEAN



- Forward - backward *like* structure
  - Forward step (major cycle) in the gradient direction of the ℓ_2 norm of the residual image
  - Backward step (minor cycle) with non-linear sparsity-enforcing operator $\mathcal{T}$

Traditional imaging techniques

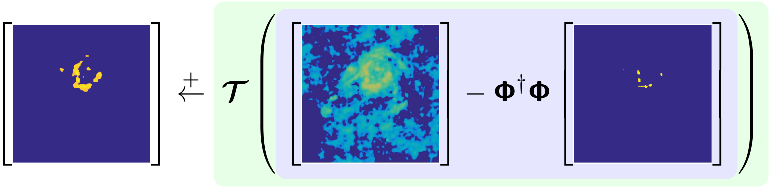
CLEAN



- Forward - backward *like* structure
  - Forward step (major cycle) in the gradient direction of the ℓ_2 norm of the residual image
  - Backward step (minor cycle) with non-linear sparsity-enforcing operator $\mathcal{T}$

Traditional imaging techniques

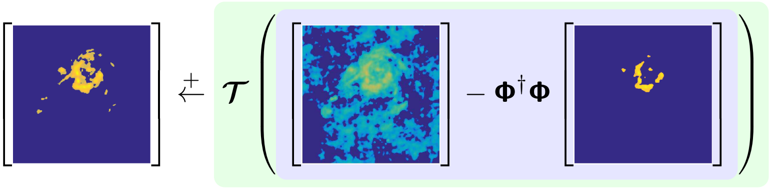
CLEAN



- Forward - backward *like* structure
  - Forward step (major cycle) in the gradient direction of the ℓ_2 norm of the residual image
  - Backward step (minor cycle) with non-linear sparsity-enforcing operator $\mathcal{T}$

Traditional imaging techniques

CLEAN



- Forward - backward *like* structure
  - Forward step (major cycle) in the gradient direction of the ℓ_2 norm of the residual image
  - Backward step (minor cycle) with non-linear sparsity-enforcing operator $\mathcal{T}$

Traditional imaging techniques

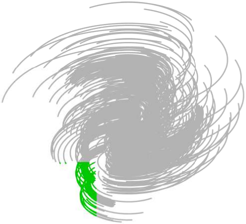
CLEAN limitations

- Sub-optimal imaging quality & manual intervention
  - Convex optimisation can deal with significantly more complex signal models in automatic mode
- Not scalable to SKA data sizes
  - Scalability of imaging methods should leverage
    - Fast transform operators modelling both data and image
    - Data Dimensionality reduction
    - Parallel and distributed processing

Data splitting

Problem formulation (1)

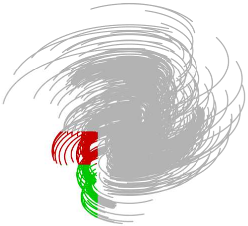
$$\begin{bmatrix} \boldsymbol{y}_1 \\ \\ \\ \end{bmatrix} =$$



Data splitting

Problem formulation (1)

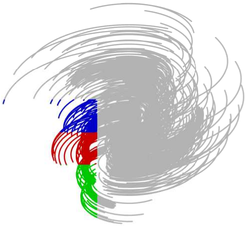
$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \\ \\ \end{bmatrix} =$$



Data splitting

Problem formulation (1)

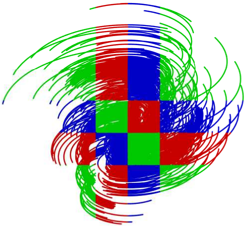
$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \boldsymbol{y}_3 \\ \\ \\ \end{bmatrix} =$$



Data splitting

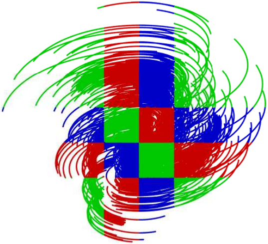
Problem formulation (1)

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \vdots \\ \mathbf{y}_d \end{bmatrix} =$$

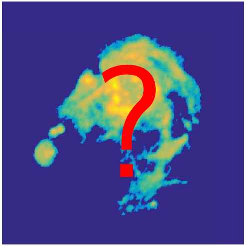


Data splitting

Problem formulation (1)



$$= \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \mathbf{G}_3 \\ \vdots \\ \mathbf{G}_d \end{bmatrix} \mathbf{F}$$

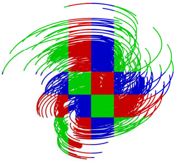


$$+ \begin{bmatrix} \boldsymbol{n}_1 \\ \boldsymbol{n}_2 \\ \boldsymbol{n}_3 \\ \vdots \\ \boldsymbol{n}_d \end{bmatrix}$$

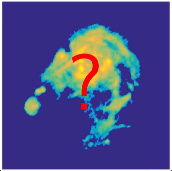
- Huge number of visibilities $\boldsymbol{y}$
  - Distribute and process the blocks independently in parallel

Data splitting

Problem formulation (1)



$$= \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \mathbf{G}_3 \\ \vdots \\ \mathbf{G}_d \end{bmatrix} \mathbf{F} \left[\quad \right] + \begin{bmatrix} \boldsymbol{n}_1 \\ \boldsymbol{n}_2 \\ \boldsymbol{n}_3 \\ \vdots \\ \boldsymbol{n}_d \end{bmatrix}$$



Distributed processing nodes

CPU $\mathbf{G}_1$ — $\boldsymbol{y}_1 =$ 

CPU $\mathbf{G}_2$ — $\boldsymbol{y}_2 =$ 

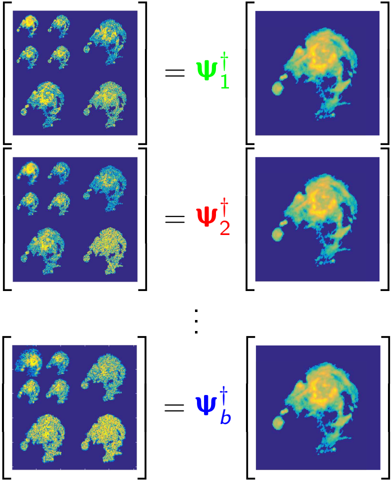
CPU $\mathbf{G}_3$ — $\boldsymbol{y}_3 =$ 

$\cdots$

CPU $\mathbf{G}_d$ — $\boldsymbol{y}_d =$ 

Enforcing sparsity priors

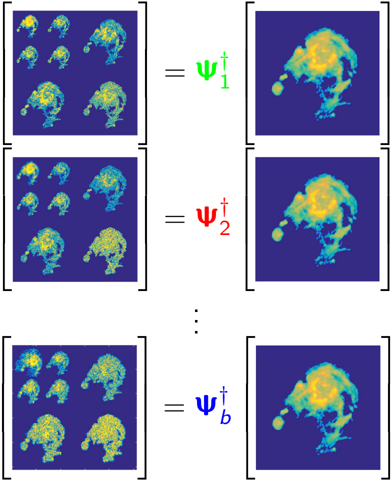
Problem formulation (2)

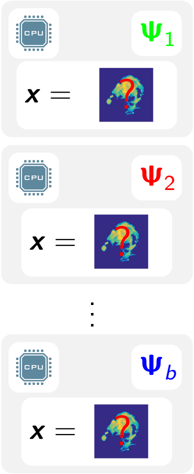


E.g. Average sparsity - a collection of wavelet bases to regularise the ill-posed problem, way beyond CLEAN.

Enforcing sparsity priors

Problem formulation (2)





Distributed imaging

Problem formulation (3)

- Split the large-scale inverse problem block wise

$$\boldsymbol{y}_j = \boldsymbol{\Phi}_j \boldsymbol{x} + \boldsymbol{n}_j \quad \text{with} \quad \boldsymbol{\Phi}_j = \mathbf{G}_j \mathbf{F}$$

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_d \end{bmatrix} \qquad \boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_1 \\ \vdots \\ \boldsymbol{\Phi}_d \end{bmatrix}$$

- Regularisation of the ill-posed problem
  - Sparsity constraint for $\boldsymbol{x}$ in a collection of wavelet bases

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_1 & \dots & \boldsymbol{\Psi}_b \end{bmatrix}$$

Distributed imaging

Problem formulation (4)

- Convex optimisation task

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \gamma \sum_{i=1}^{b} l_i(\boldsymbol{\Psi}_i^\dagger \boldsymbol{x}) + \sum_{j=1}^{d} h_j(\boldsymbol{\Phi}_j \boldsymbol{x})$$

- Enforce positivity, sparsity and data fidelity

$$f(\boldsymbol{z}) = \iota_{\mathcal{C}}(\boldsymbol{z}), \mathcal{C} = \mathbb{R}_+^N$$

$$l_i(\boldsymbol{z}) = \|\boldsymbol{z}\|_1$$

$$h_j(\boldsymbol{z}) = \iota_{\mathcal{B}_j}(\boldsymbol{z}), \;\; \mathcal{B}_j = \{\boldsymbol{z} \in \mathbb{C}^{M_j} : \|\boldsymbol{z} - \boldsymbol{y}_j\|_2 \leq \epsilon_j\}$$

Distributed imaging

The primal dual approach

- Primal problem

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \gamma \sum_{i=1}^{b} l_i(\boldsymbol{\Psi}_i^{\dagger} \boldsymbol{x}) + \sum_{j=1}^{d} h_j(\boldsymbol{\Phi}_j \boldsymbol{x})$$

- Dual formulation of the original convex optimisation task

$$\min_{\substack{\boldsymbol{u}_i \\ \boldsymbol{v}_j}} f^* \left(- \sum_{i=1}^{b} \boldsymbol{\Psi}_i \boldsymbol{u}_i - \sum_{j=1}^{d} \boldsymbol{\Phi}_j^{\dagger} \boldsymbol{v}_j \right) + \frac{1}{\gamma} \sum_{i=1}^{b} l_i^*(\boldsymbol{u}_i) + \sum_{j=1}^{d} h_j^*(\boldsymbol{v}_j)$$

- Primal dual algorithm
  - Alternate solving the primal problem and the dual problem
  - Converges towards a Kuhn-Tucker point

Distributed imaging

Advantages of the primal dual approach

- Full splitting of the operators and functions
- No inversion of the linear operators
- Forward-backward iterations applied in parallel for all dual variables in data, sparsity, and image space
  - Interlaced and parallel CLEAN-like iteration structure
- Randomised updates on the dual variables to reduce computational and memory needs per iteration
- Non-Euclidean updates on the dual variables to accelerate convergence

Primal dual algorithm

given $\boldsymbol{x}^{(0)}, \tilde{\boldsymbol{x}}^{(0)}, \boldsymbol{u}_i^{(0)}, \boldsymbol{v}_j^{(0)}, \tilde{\boldsymbol{u}}_i^{(0)}, \tilde{\boldsymbol{v}}_j^{(0)}, \gamma, \tau, \sigma_i$

repeat for $t = 1, \ldots$

generate sets $\mathcal{P} \subset \{1, \ldots, b\}$ **and** $\mathcal{D} \subset \{1, \ldots, d\}$

$$\boldsymbol{b}_j^{(t)} = \mathbf{M}_j \mathbf{F} \mathbf{Z} \tilde{\boldsymbol{x}}^{(t-1)}, \quad \forall j \in \mathcal{D}$$

run simultaneously

$\forall j \in \mathcal{D}$ **distribute** $\boldsymbol{b}_j^{(t)}$ **and do in parallel**

$$\boldsymbol{v}_j^{(t)} = \left(\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{P}}_{\mathcal{B}_j}\right)\left(\boldsymbol{v}_j^{(t-1)} + \mathbf{G}_j \boldsymbol{b}_j^{(t)}\right)$$

$$\tilde{\boldsymbol{v}}_j^{(t)} = \mathbf{G}_j^* \boldsymbol{v}_i^{(t)}$$

"CLEAN $\tilde{\boldsymbol{v}}_j$"

end and gather $\tilde{\boldsymbol{v}}_j^{(t)}$

$\forall i \in \mathcal{P}$ **do in parallel**

$$\boldsymbol{u}_i^{(t)} = \left(\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{S}}_{\frac{\gamma}{\sigma_i}}\right)\left(\boldsymbol{u}_i^{(t-1)} + \boldsymbol{\Psi}_i^* \tilde{\boldsymbol{x}}^{(t)}\right)$$

$$\tilde{\boldsymbol{u}}_i^{(t)} = \boldsymbol{\Psi}_i \boldsymbol{u}_i^{(t)}$$

"CLEAN $\tilde{\boldsymbol{u}}_i$"

end

end

$$\bar{\boldsymbol{x}}^{(t)} = \boldsymbol{\mathcal{P}}_{\mathcal{C}}\left(\boldsymbol{x}^{(t-1)} - \tau\left(\sum_{i=1}^{b} \sigma_i \tilde{\boldsymbol{u}}_i^{(t)} + \mathbf{Z}^* \mathbf{F}^\dagger \sum_{j=1}^{d} \varsigma_j \mathbf{M}_j^* \tilde{\boldsymbol{v}}_j^{(t)}\right)\right)$$

"CLEAN $\tilde{\boldsymbol{x}}$"

$$\tilde{\boldsymbol{x}}^{(t)} = 2\bar{\boldsymbol{x}}^{(t)} - \boldsymbol{x}^{(t-1)}$$

until convergence

The story of a black hole discovery?

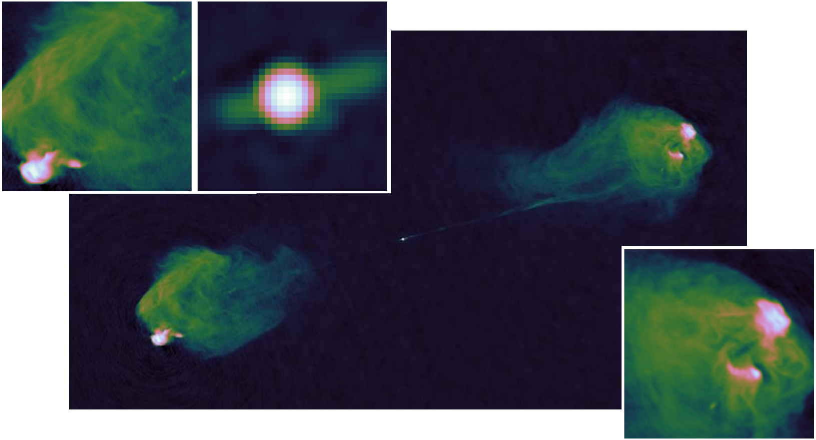
Imaging Cygnus A Galaxy from real data

- NRAO recently confirmed discovery of a new source, just 460 pc from the nucleus of Cygnus A, likely to be a second black hole. Discovery from JVLA data using CLEAN at X-band (8.5 GHz). *[official optimised pipeline]*

- My team, in collaboration with NRAO and SKA South Africa, is confirming discovery from JVLA data at C-band (6.6GHz) using our convex optimisation algorithm, where CLEAN is blind to this angular resolution. *[home-made MATLAB solver]*

The story of a black hole discovery?

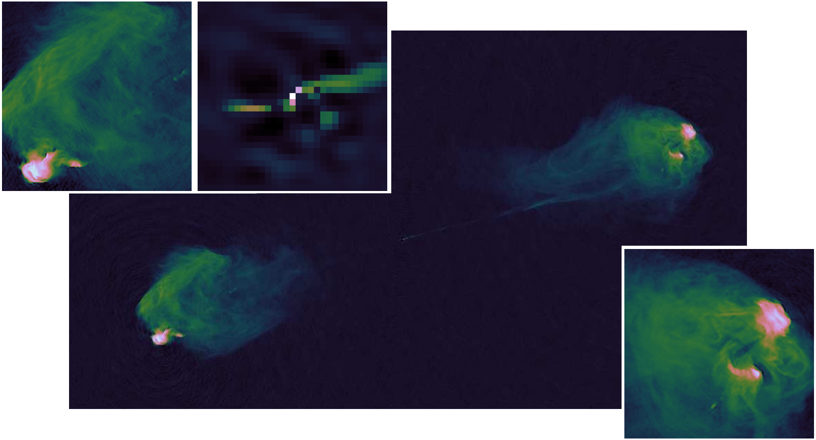
CLEAN results at C-band (6.6GHz) - 16GB image



The story of a black hole discovery?

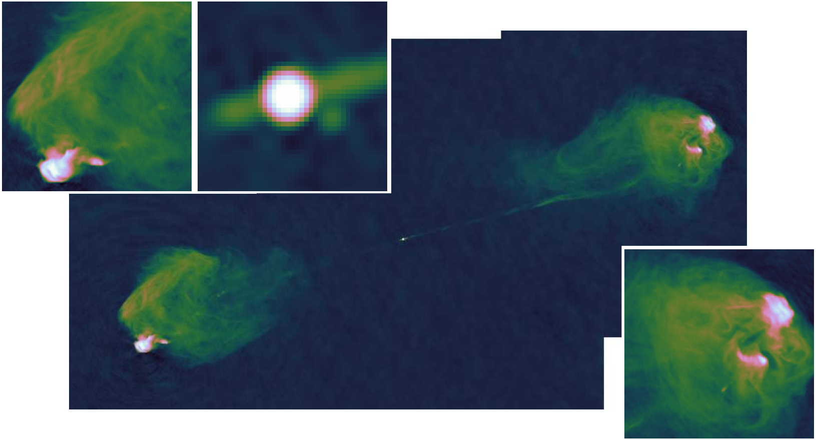
Primal Dual results at C-band (6.6GHz) - 16GB image

- Super-resolution: restored image at 2.1× the resolution of the instrument.



The story of a black hole discovery?

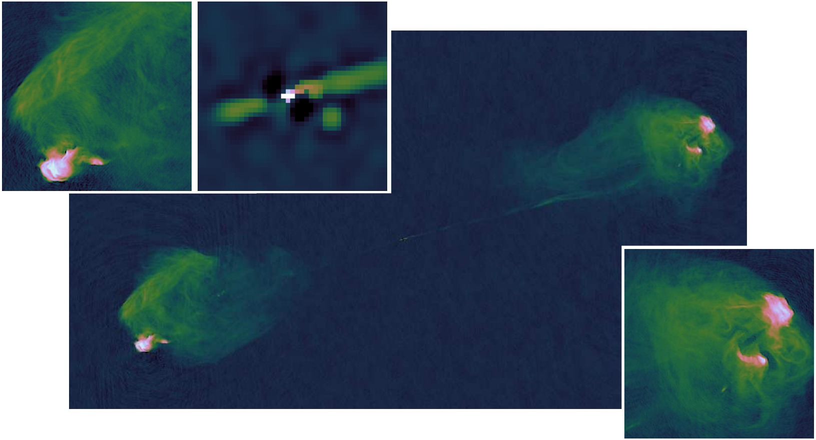
CLEAN results at X-band (8.5GHz) - 16GB image



The story of a black hole discovery?

Primal Dual results at X-band (8.5GHz) - 16GB image

- Super-resolution: restored image at 2.5× the resolution of the instrument.



Concluding with new project award

A 4 year project was recently funded jointly by the Swiss NSF & South African NRF

- Partners :
  - Edinburgh: Prof Wiaux
  - SKA South Africa: Prof Smirnov (with Profs Gain & Cress)
  - EPFL: Prof Thiran (with Profs Vandergheynst & Kneib)
- Researchers :
  - EPFL - Edinburgh: 1 Postdoctoral researcher (3 years)
  - SKA South Africa: 2 PhD Students (4 years)
- Science :
  - Scalable Algorithms for hyperspectral and polarisation imaging
  - HPC implementation
  - Validation from Meerkat data
  - New discoveries...