Primal-dual algorithms for next-generation radio-interferometric imaging

Alexandru Onose[1], Rafael E. Carrillo[2], Abdullah Abdulaziz[1], Arwa Dabbech[1], Jason D. McEwen[3] and Yves Wiaux[1]

[1]Institute of Sensors, Signals and Systems, Heriot-Watt University, United Kingdom
[2]Signal Processing Laboratory, Ecole Polytechnique Fédérale de Lausanne, Switzerland
[3]Mullard Space Science Laboratory, University College London, United Kingdom

Next-generation radio-interferometric challenges

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

gigapixel images

huge dynamic range

Next-generation radio-interferometric challenges

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

  gigapixel images   huge dynamic range

- Unprecedented amount of data to be processed
- Good reconstruction quality with scalable algorithms employing parallel and distributed processing

Measurement model

- 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 inverse problem

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

  - $\boldsymbol{x} \in \mathbb{R}_+^N$ 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
  - $\mathbf{FZ} \in \mathbb{C}^{kN\times N}$ Fourier matrix with zero padding

Measurement model

- 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 inverse problem

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

  - $\boldsymbol{x} \in \mathbb{R}_+^N$ 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
  - $\mathbf{FZ} \in \mathbb{C}^{kN\times N}$ Fourier matrix with zero padding

extremely large M and N

Measurement model

- 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 inverse problem

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

  - $\boldsymbol{x} \in \mathbb{R}_+^N$ 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
  - $\mathbf{FZ} \in \mathbb{C}^{kN \times N}$ Fourier matrix with zero padding

extremely large M and N

scalable algorithms; parallelisms and distributed processing

Convex optimisation and compressive sensing

Problem formulation (0)

- Discretised version of the ill-posed inverse problem

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

- Compressive sensing approach
  - Constrain the solution to belong to the intersection of several convex sets enforcing data fidelity and positivity
  - Add a convex regulariser on the solution to enforce low dimensionality, sparsity, in a given transform
- Convex optimisation solvers
  - Primal-dual forward-backward algorithm
    - Distributed single-band RI imaging
    - Wide-band RI imaging

Distributed single-band RI imaging

Problem formulation (1)

- Split the large-scale inverse problem block wise

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

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

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

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_1 & \ldots & \boldsymbol{\Psi}_{n_\mathrm{b}} \end{bmatrix}$$

Distributed single-band RI imaging

Problem formulation (2)

- Convex optimisation task

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

- 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 single-band RI imaging

The primal dual approach

- Primal problem

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

- Dual formulation of the original convex optimisation task

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

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

Distributed single-band RI imaging

Advantages of the primal dual approach

- Full splitting of the operators and functions
  - No inversion of the linear operators
  - The updates are performed on the dual variables in parallel
- Non smooth convex functions
  - Forward backward iterations
  - Alternate between a gradient like step (forward) and a proximal update (backward)
- Randomised updates on the dual variables
  - Reduce computational and memory requirements per iteration
  - Require more iterations to converge

Distributed single-band RI imaging

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 $\mathcal{P} \subset \{1, \ldots, n_{\mathrm{b}}\}$ **and** $\mathcal{D} \subset \{1, \ldots, n_{\mathrm{d}}\}$

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

run simultaneously

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

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

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

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

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

end

end

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

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

until convergence

Distributed single-band RI imaging

Computational complexity

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

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

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

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

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

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

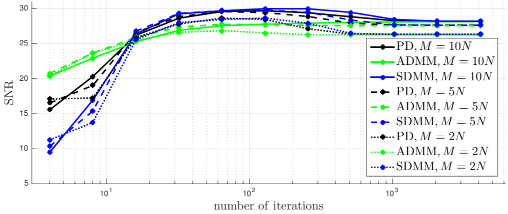
end

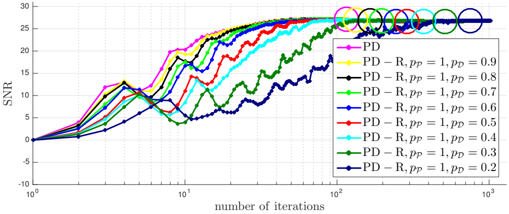
$$\bar{\boldsymbol{x}}^{(t)} = \boldsymbol{\mathcal{P}}_{\mathcal{C}}\left(\ldots\ldots\ldots\ldots\right)$$

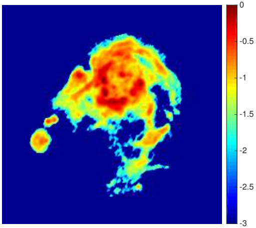
central node	n_{d} data fidelity nodes
$\mathcal{O}\left(kN \log kN\right)$	—
$p_{\mathcal{P}_i}\mathcal{O}\left(2n_{\mathrm{b}}N\right)$	$p_{\mathcal{D}_j}\mathcal{O}\left(2\frac{kn_{\mathrm{s}}}{n_{\mathrm{d}}}MN_j + M_j\right)$
$\mathcal{O}\left(kN \log kN\right) + \mathcal{O}\left((n_{\mathrm{b}} + k)N + n_{\mathrm{d}}n_{\mathrm{v}}\right)$	—

Distributed single-band RI imaging

Some results



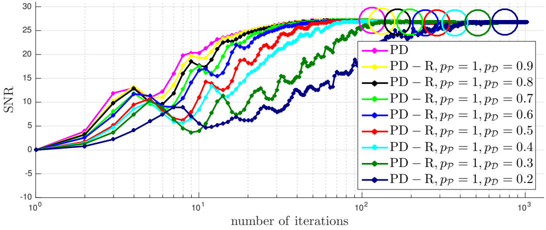


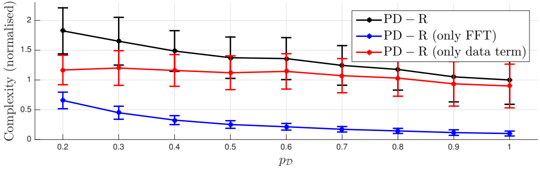


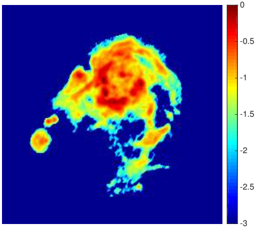
- R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "PURIFY: a new approach to radio-interferometric imaging," MNRAS, vol. 439, no. 4, pp. 3591-3604, 2014.

Distributed single-band RI imaging

Some results







Wide-band RI imaging

Problem formulation (0)

- Multiple frequency bands; images $\boldsymbol{x}_l$ at each band $l = 1, \ldots, L$
- Discretised version of the ill-posed inverse problem

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

  - $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L] \in \mathbb{R}_+^{N \times L}$ the hyper-spectral image cube
  - $\mathbf{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_L] \in \mathbb{C}^{M \times L}$ the measured visibilities
  - $\boldsymbol{\Phi}_l \in \mathbb{C}^{M \times N}$ a linear map; image domain to visibility space
- Compressive sensing approach
  - Enforce data fidelity and positivity
  - Convex regulariser to enforce low dimensionality
    - Joint sparsity and low rank on $\mathbf{X}$
- Primal-dual forward-backward algorithm

Wide-band RI imaging

Problem formulation (1)

- Minimisation problem

$$\min_{\mathbf{X}} f(\mathbf{X}) + \mu\, g_1(\mathbf{\Psi}^\dagger \mathbf{X}) + g_2(\mathbf{X}) + \sum_{l=1}^{L} h_l(\mathbf{\Phi}_l \boldsymbol{x}_l))$$

- Enforce positivity, joint sparsity, low rank, data fidelity

$$f(\mathbf{Z}) = \iota_{\mathcal{D}}(\mathbf{Z}), \ \ \mathcal{D} = \mathbb{R}_+^{N \times L}$$

$$g_1(\mathbf{Z}) = \|\mathbf{Z}\|_{\ell_{2,1}}, \quad g_2(\mathbf{Z}) = \|\mathbf{Z}\|_*$$

$$h_l(\boldsymbol{z}) = \iota_{\mathcal{B}_l}(\boldsymbol{z}), \ \ \mathcal{B}_l = \{\boldsymbol{z} \in \mathbb{C}^M : \|\boldsymbol{z} - \boldsymbol{y}_l\|_2 \leq \epsilon_l\}$$

Wide-band RI imaging

Primal dual algorithm

given $\mathbf{X}^{(0)}, \tilde{\mathbf{X}}^{(0)}, \mathbf{V}_1^{(0)}, \mathbf{V}_2^{(0)}, \mathbf{V}_3^{(0)}, \mu, \tau, \sigma_1, \sigma_2, \sigma_3$
repeat for $t = 1, \ldots$
do in parallel

$$\mathbf{V}_1^{(t)} = \mathbf{V}_1^{(t-1)} + \boldsymbol{\Psi}^\dagger \tilde{\mathbf{X}}^{(t-1)} - \mathcal{S}_{\mu/\sigma_1}^{\ell_{2,1}}\left(\mathbf{V}_1^{(t-1)} + \boldsymbol{\Psi}^\dagger \tilde{\mathbf{X}}^{(t-1)}\right)$$

$$\mathbf{V}_2^{(t)} = \mathbf{V}_2^{(t-1)} + \tilde{\mathbf{X}}^{(t-1)} - \mathcal{S}_{1/\sigma_2}^{*}\left(\mathbf{V}_2^{(t-1)} + \tilde{\mathbf{X}}^{(t-1)}\right)$$

$\forall l \in \{1, \ldots, L\}$ **do in parallel**

$$\mathbf{u}_l^{(t)} = \mathbf{u}_l^{(t-1)} + \boldsymbol{\Phi}_l \tilde{\mathbf{x}}_l^{(t-1)} - \mathcal{P}_{\mathcal{B}_l}\left(\boldsymbol{u}_l^{(t-1)} + \boldsymbol{\Phi}_l \tilde{\mathbf{x}}_l^{(t-1)}\right)$$

end
end

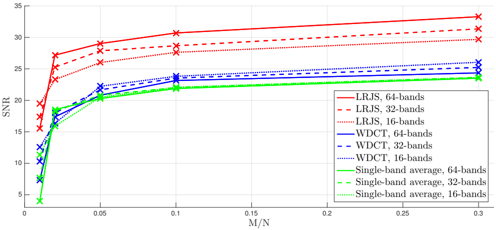
$$\mathbf{X}^{(t)} = \mathcal{P}_{\mathcal{C}}\left(\mathbf{X}^{(t-1)} - \tau\left(\sigma_1 \boldsymbol{\Psi} \mathbf{V}_1^{(t)} + \sigma_2 \mathbf{V}_2^{(t)} + \sigma_3 \left[\boldsymbol{\Phi}_1^\dagger \boldsymbol{u}_1^{(t)}, \ldots, \boldsymbol{\Phi}_L^\dagger \boldsymbol{u}_L^{(t)}\right]\right)\right)$$

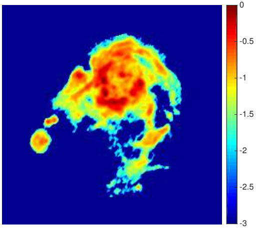
$$\tilde{\mathbf{X}}^{(t)} = 2\mathbf{X}^{(t)} - \mathbf{X}^{(t-1)}$$

until convergence

Distributed single-band RI imaging

Some results





- A. Ferrari, J. Deguignet, C. Ferrari, D. Mary, A. Schutz, and O. Smirnov, "Multi-frequency image reconstruction for radio interferometry. A regularized inverse problem approach," ArXiv e-prints, Apr. 2015.

Summary

- Full splitting of the operators and functions
  - No inversion of the linear operators
- Broad range of convex prior functions
  - Forward backward iterations for non smooth functions
- Randomisation
  - Reduce computational and memory requirements
  - Require more iterations to converge