PURIFY: a new algorithmic framework for next-generation radio-interferometric imaging

Rafael E. Carrillo

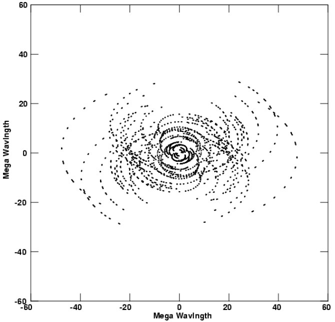
Signal Processing Laboratory (LTS5)
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
&
BASP Group
Heriot-Watt University, UK

Work in collaboration with J. D. McEwen and Y. Wiaux

CALIM March 5 2014

Introduction

Interferometers provide incomplete Fourier measurements of the observed object (complex visibilities)



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

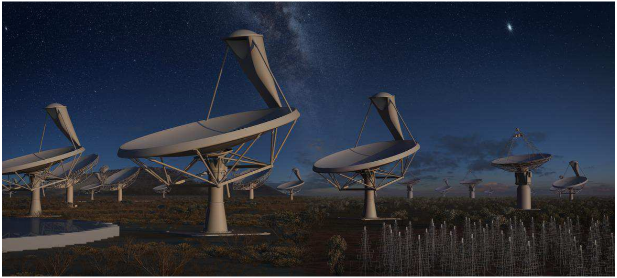
- $A(\mathbf{l}, \mathbf{u})$: direction dependent effects

Image recovery poses a linear inverse problem:

$$\mathbf{y} = \Phi\mathbf{x}, \text{ with } \Phi \in \mathbb{C}^{M \times N}$$

Introduction

Next generation telescopes, such as the SKA, has triggered an intense research to reformulate imaging techniques for radio interferometry.



Motivation

Main challenges for next generation telescopes

- High resolution and dynamic range
- Large number of visibilities ($M \approx 10^6 N$)

Motivation

Main challenges for next generation telescopes

- High resolution and dynamic range
- Large number of visibilities ($M \approx 10^6 N$)

Our solution

- Leverage recent advances in compressed sensing (CS) and convex optimization to address these challenging problems
- Effectiveness of compressed sensing applied to radio interferometric imaging already demonstrated (Wiaux et al. 2009a, Wiaux et al. 2009b, McEwen & Wiaux 2011, Li et al. 2011, Carrillo et al. 2012)

Outline

Outline

CS Signal Recovery (I)

- Suppose $\mathbf{x}$ is expressed in terms of a basis $\Psi \in \mathbb{R}^{N \times N}$, as $\mathbf{x} = \Psi \boldsymbol{\alpha}$, $\boldsymbol{\alpha} \in \mathbb{R}^N$
- Noisy model:

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

- Two different approaches
  - Synthesis based problem:

$$\min_{\bar{\boldsymbol{\alpha}} \in \mathbb{R}^N} \|\bar{\boldsymbol{\alpha}}\|_1 \text{ subject to } \|\mathbf{y} - \Phi \Psi \bar{\boldsymbol{\alpha}}\|_2 \leq \epsilon$$

  - Analysis based problem:

$$\min_{\bar{\mathbf{x}} \in \mathbb{R}^N} \|\Psi^\dagger \bar{\mathbf{x}}\|_1 \text{ subject to } \|\mathbf{y} - \Phi \bar{\mathbf{x}}\|_2 \leq \epsilon$$

CS Signal Recovery (II)

- Most CS approaches solve the Lagrangian formulation:

$$\min_{\bar{\boldsymbol{\alpha}} \in \mathbb{C}^N} \frac{1}{2}\|\mathbf{y} - \Phi\Psi\bar{\boldsymbol{\alpha}}\|_2^2 + \lambda\|\bar{\boldsymbol{\alpha}}\|_1$$

- Update equation:

$$\boldsymbol{\alpha}^{(t+1)} = \mathsf{S}_\lambda\left(\boldsymbol{\alpha}^{(t)} + \mu\Psi^\dagger\Phi^\dagger(\mathbf{y} - \Phi\Psi\boldsymbol{\alpha}^{(t)})\right)$$

- Efficient algorithms to solve this problem such as FISTA (Beck and Teboulle 2009)
- However there is no optimal strategy to estimate λ

Average Sparsity

- We recently propose the SARA algorithm based on the average sparsity model
- It uses a dictionary composed of several coherent frames:

$$\Psi = [\Psi_1, \Psi_2, \ldots, \Psi_q]$$

- Optimization problem:

$$\min_{\bar{\mathbf{x}} \in \mathbb{R}_+^N} \|\Psi^\dagger \bar{\mathbf{x}}\|_0 \text{ subject to } \|\mathbf{y} - \Phi \bar{\mathbf{x}}\|_2 \leq \epsilon$$

$$\|\Psi^\dagger \bar{\mathbf{x}}\|_0 = \sum_{i=1}^{q} \|\Psi_i^\dagger \bar{\mathbf{x}}\|_0 \rightarrow \text{average sparsity}$$

- A reweighting scheme solving a sequence of (convex) weighted ℓ_1-problems is used to approximate the ℓ_0 problem

Constrained Optimization

Thus we focus on solving problems of the form:

$$\min_{\bar{\mathbf{x}} \in \mathbb{R}_+^N} \|\mathsf{W}\Psi^\dagger \bar{\mathbf{x}}\|_1 \text{ subject to } \|\mathbf{y} - \Phi\bar{\mathbf{x}}\|_2 \leq \epsilon$$

- $\epsilon = \sigma_n \sqrt{M + 2\sqrt{M}}$ $\rightarrow$statistical bound
- $\bar{\mathbf{x}} \in \mathbb{R}_+^N$ $\rightarrow$positivity constraint
- $\Phi = \mathsf{GFDA}$
  - G : convolutional interpolation operator
  - F : fast Fourier transform
  - D : deconvolution operator
  - A : primary beam

Outline

A Large-scale Optimization Algorithm

- Large-scale data problems, i.e. $M \gg N$ and large N
- Partition $\mathbf{y}$ and Φ into R blocks:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_R \end{bmatrix} \text{ and } \Phi = \begin{bmatrix} \Phi_1 \\ \vdots \\ \Phi_R \end{bmatrix}$$

- Each $\mathbf{y}_i$ is modeled as $\mathbf{y}_i = \Phi_i \mathbf{x} + \mathbf{n}_i$
- Reconstruction problem reformulated as

$$\min_{\bar{\mathbf{x}} \in \mathbb{R}_+^N} \|W\Psi^\dagger \bar{\mathbf{x}}\|_1 \text{ subject to } \|\mathbf{y}_i - \Phi_i \bar{\mathbf{x}}\|_2 \leq \epsilon_i, i = 1, \ldots, R$$

Proximal Splitting Methods

- Solve problems of the form

$$\min_{\mathbf{x}\in\mathbb{R}^N} f_1(\mathbf{x}) + \ldots + f_S(\mathbf{x})$$

- $f_1(\mathbf{x}), \ldots, f_S(\mathbf{x})$ are proper convex lower semicontinuous functions from $\mathbb{R}^N$ to $\mathbb{R}$ (not necessarily differentiable)
- Key idea: split a complicated problem into several simpler problems
- Each non-smooth function is incorporated in the optimization via its proximity operator:

$$\mathrm{prox}_f(\mathbf{x}) \triangleq \arg\min_{\mathbf{z}\in\mathbb{R}^N} f(\mathbf{z}) + \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2$$

Projection onto Convex Sets

- Proximity operators are generalizations of the set projection operator

$$P_C(\mathbf{x}) = \arg\min_{\mathbf{z}\in C} \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2$$

- Any convex constraint $\mathbf{z} \in C$ can be modelled by its indicator function

$$i_C(\mathbf{z}) = \begin{cases} 0, & \text{if } \mathbf{z} \in C \\ +\infty, & \text{otherwise} \end{cases}$$

- Proximity operator of indicator function

$$\begin{aligned} P_C(\mathbf{x}) &= \arg\min_{\mathbf{z}\in\mathbb{R}^N} i_C(\mathbf{z}) + \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2 \\ &= \mathrm{prox}_{i_C}(\mathbf{x}) \end{aligned}$$

Solving the Weighted ℓ_1 Problem

The ℓ_1 problem can be reformulated as:

$$\min_{\mathbf{x} \in \mathbb{R}^N} f_1(\mathsf{L}_1 \mathbf{x}) + \cdots + f_S(\mathsf{L}_S \mathbf{x})$$

with $S = R + 2$

- $\mathsf{L}_1 = \Psi^\dagger$, $\mathsf{L}_2 = \mathsf{I}$ and $\mathsf{L}_{k+2} = \Phi_k$ for $k = 1, \ldots, S$
- $f_1(\mathbf{r}_1) = \|\mathsf{W}\mathbf{r}_1\|_1$ for $\mathbf{r}_1 \in \mathbb{R}^D$
- $f_2(\mathbf{r}_2) = i_C(\mathbf{r}_2)$ with $C = \mathbb{R}^N_+$
- $f_k(\mathbf{r}_k) = i_{B_k}(\mathbf{r}_k)$ with $B_k = \{\mathbf{r}_k \in \mathbb{R}^{M_k} : \|\mathbf{y_k} - \mathbf{r}_k\|_2 \leq \epsilon_k\}$, $k = 3, \ldots, S$

Simultaneous-Direction Method of Multipliers (SDMM)

SDMM uses the following equivalent problem

$$\min f_1(\mathbf{r}_1) + \ldots + f_S(\mathbf{r}_3)$$
$$\text{subject to } \mathsf{L}_k \mathbf{x} = \mathbf{r}_k, \text{ for } k = 1, \ldots, S$$

- SDMM decouples the problems for $f_1, \ldots, f_S$
- Subproblems optimizing $f_1, \ldots, f_S$ no longer involve linear operators
- Optimization based in an alternate primal-dual approach

SDMM Algorithm

Initialize $\gamma > 0$, $\hat{\mathbf{x}}^{(0)}$ and $\mathbf{z}_i^{(0)} = \mathbf{0}$, $i = 1, \ldots, S$.
$\mathbf{r}_i^{(0)} = \mathsf{L}_i \hat{\mathbf{x}}^{(0)}$, $i = 1, \ldots, S$.
$\mathbf{x}_i^{(0)} = \mathsf{L}_i^{\dagger} \mathbf{r}_i^{(0)}$, $i = 1, \ldots, S$.
while No convergence criteria **do**
  $\hat{\mathbf{x}}^{(t)} = (\sum_{i=1}^{S} \mathsf{L}_i^{\dagger} \mathsf{L}_i)^{-1} \sum_{i=1}^{S} \mathbf{x}_i^{(t-1)}$.
  for all $i = 1, \ldots, S$ **do**
    $\mathbf{r}_i^{(t)} = \mathrm{prox}_{\gamma f_i}(\mathsf{L}_i \hat{\mathbf{x}}^{(t)} + \mathbf{z}_i^{(t-1)})$.
    $\mathbf{z}_i^{(t)} = \mathbf{z}_i^{(t-1)} + \mathsf{L}_i \hat{\mathbf{x}}^{(t)} - \mathbf{r}_i^{(t)}$.
    $\mathbf{x}_i^{(t)} = \mathsf{L}_i^{\dagger} (\mathbf{r}_i^{(t)} - \mathbf{z}_i^{(t)})$.
  end for
end while
return $\hat{\mathbf{x}}^{(t)}$

Implementation Details

Proximal operators

$$\mathrm{prox}_{\gamma f_1}(\mathbf{r}_1) = S_\gamma(\mathbf{r}_1) \rightarrow \text{soft thresholding}$$
$$\mathrm{prox}_{\gamma f_2}(\mathbf{r}_2) = (\mathbf{r}_2)^+ \rightarrow \text{thresholding of negative values}$$
$$\mathrm{prox}_{\gamma f_k}(\mathbf{r}_k) = \min(1, \epsilon_k / \|\mathbf{r}_k\|_2)\mathbf{r}_k \rightarrow \text{scaling} \ , k = 3, \ldots, S$$

- Very simple element wise operations
- Can be performed in parallel!

Implementation Details

Linear system

$$\mathbf{x}^{(t)} = \Big(\sum_{i=1}^{S} \mathsf{L}_i^{\dagger}\mathsf{L}_i\Big)^{-1} \sum_{i=1}^{S} \mathsf{L}_i^{\dagger}\big(\mathbf{r}_i^{(t-1)} - \mathbf{z}_i^{(t-1)}\big)$$

- Solved iteratively using a conjugate gradient algorithm
- For the problem in hand $\sum_{i=1}^{S} \mathsf{L}_i^{\dagger}\mathsf{L}_i = \Phi^{\dagger}\Phi + 2\mathsf{I}$
- Bottleneck of the algorithm!
- Matrix inversion lemma can be used to accelerate the inversion of $\Phi^{\dagger}\Phi + 2\mathsf{I}$

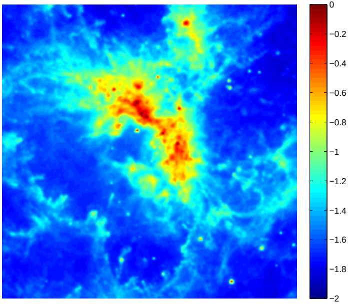
PURIFY

- PURIFY is an open-source code that provides functionality to perform radio interferometric imaging
- SDMM based solvers for the optimization problems
- Implements the following sparsity priors:
  - Daubechies orthogonal wavelets
  - Total variation
  - Sparsity averaging
- Code available at github (http://basp-group.github.io/purify/)

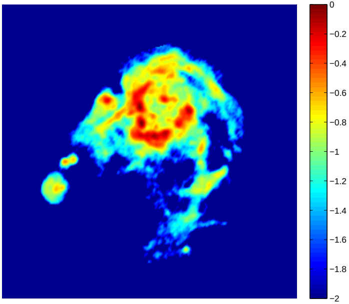
Outline

Test Images

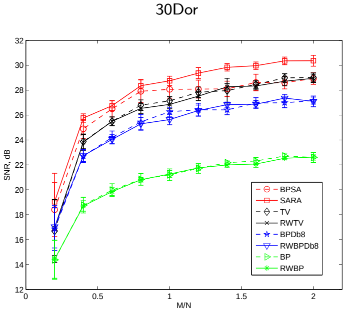
30Dor

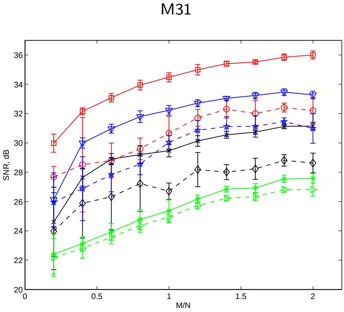


M31

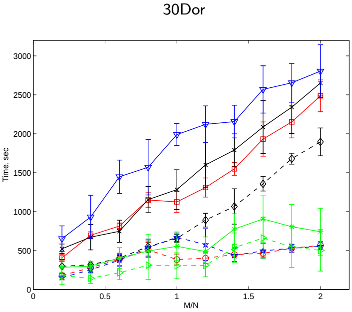


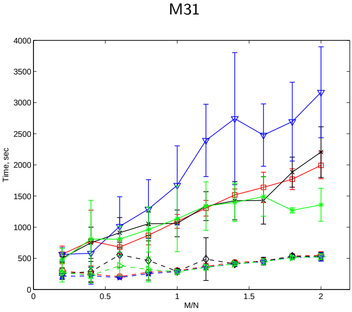
Reconstruction Quality Results





Timing Results (Not Optimized)





Outline

CS Signal Recovery

Large-scale Optimization

Numerical Results

Summary

Summary

- SARA supersedes state-of-the-art reconstruction algorithms for RI imaging
- We developed an open source code (PURIFY) to scale to the realistic setting
- Direction dependent effects can be incorporated in the model as convolutional kernels in the operator G (see next talk by Jason)
- New ways to improve the computational efficiency of the algorithm have to be explored:
  - Specialized hardware implementations
  - Distributed approaches
  - Dimensionality reduction techniques

Thank You!