Expectation Maximization Type Algorithms for Calibration

Sarod Yatawatta

Kapteyn Institute, University of Groningen

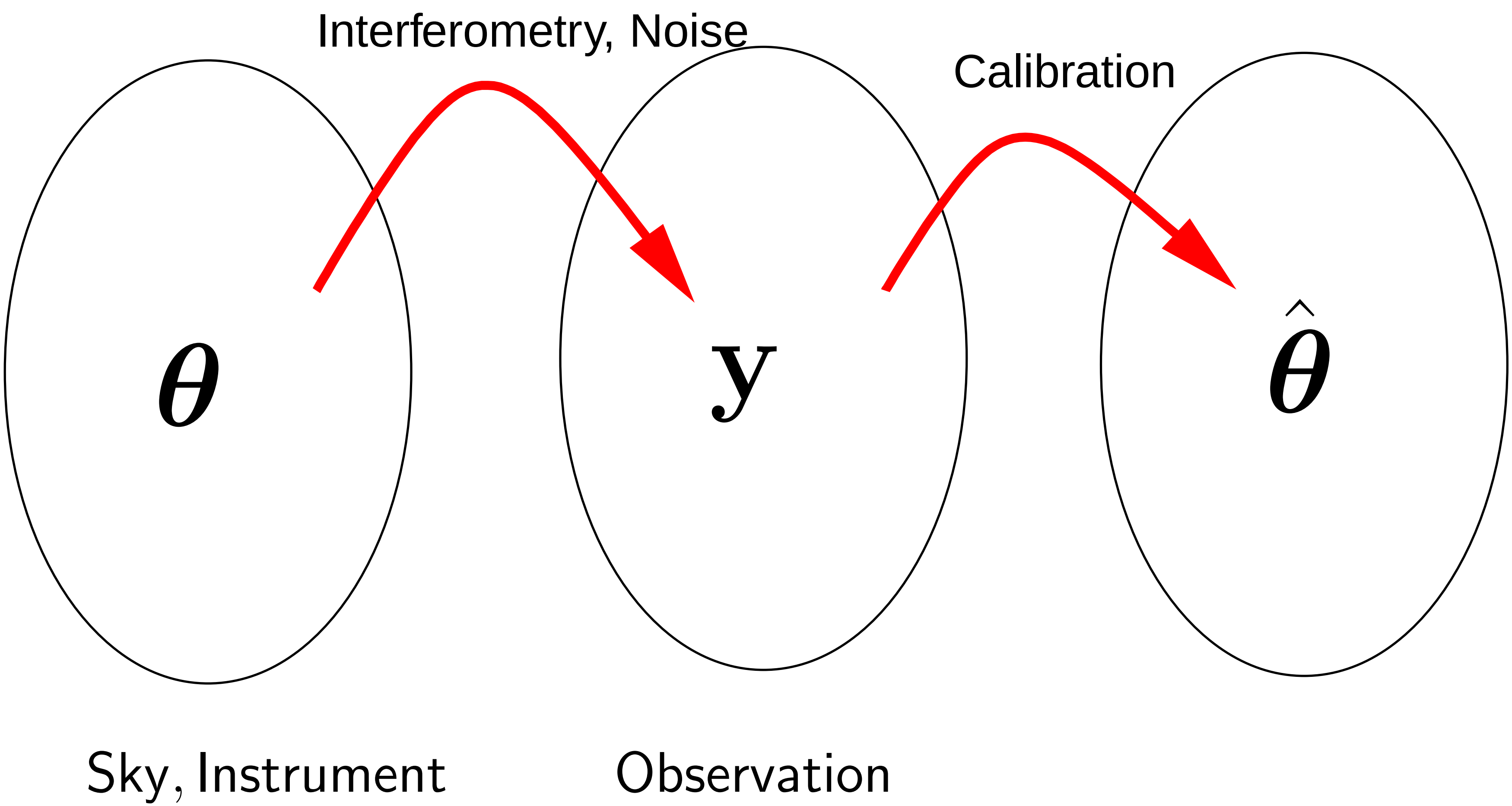
and

ASTRON

collaborators:

Saleem Zaroubi, Ger de Bruyn, Leon Koopmans and Jan Noordam

Calibration



Ideally, find $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$. But in real life ???

Calibration

For K discrete sources, we observe

$$\mathbf{y} = \sum_{i=1}^{K} \mathbf{s}_i(\boldsymbol{\theta}) + \mathbf{n}$$

Maximum Likelihood (ML) estimate, under White Gaussian Noise

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\theta}} \|\mathbf{y} - \sum_{i=1}^{K} \mathbf{s}_i(\boldsymbol{\theta})\|^2$$

Traditional calibration: using Levenberg-Marquardt (LM) algorithm

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^{k} - (\boldsymbol{\nabla}_{\boldsymbol{\theta}}\boldsymbol{\nabla}_{\boldsymbol{\theta}}^{T}\phi(\boldsymbol{\theta}) + \lambda\mathbf{H})^{-1}\boldsymbol{\nabla}_{\boldsymbol{\theta}}\phi(\boldsymbol{\theta})|_{\boldsymbol{\theta}^{k}}$$

where $\mathbf{H} \stackrel{\triangle}{=} \mathrm{diag}(\boldsymbol{\nabla}_{\boldsymbol{\theta}}\boldsymbol{\nabla}_{\boldsymbol{\theta}}^{T}\phi(\boldsymbol{\theta}))$.

LM in Action

The long and winding road

Expectation Maximization

- Formally introduced by [Dempster, Laird, Rubin, 77]
- ML estimation with incomplete data
- Iterative maximization of the likelihood
- Frequently used but most misunderstood algorithm

E.g. Gain/Leakage calibartion with an unpolarized calibrator

$$\mathbf{V}_{pq} = I\mathbf{G}_p\mathbf{D}_p\mathbf{D}_q^H\mathbf{G}_q^H + \mathbf{N}$$

$$\mathbf{G}_p = \left[\begin{array}{cc} g_{x,p} & 0 \\ 0 & g_{y,p} \end{array}\right],\ \ \mathbf{D}_p = \left[\begin{array}{cc} 1 & \delta_{x,p} \\ \delta_{y,p} & 1 \end{array}\right], \delta_{x,p}, \delta_{y,p} \ll 1$$

First Estimate $\mathbf{G}$ and then $\mathbf{D}$, and iterate.
But, does the other way work?

university of groningen ASTRON

EM: Formal Description

$$\mathbf{y} = \sum_{i=1}^{K} \mathbf{s}_i(\boldsymbol{\theta}) + \mathbf{n}$$

- ML estimate: $\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} \log f(\mathbf{y}|\boldsymbol{\theta})$
- Auxiliary random variable $\mathbf{x}$: hidden data, $\mathbf{y} = \mathbf{F}(\mathbf{x})$
- The *E Step*: compute conditional expectation $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k) = E\{\log f(\mathbf{x}|\boldsymbol{\theta})|\mathbf{y}, \boldsymbol{\theta}^k\}$
- The *M Step*: Maximize $\boldsymbol{\theta}^{k+1} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^k)$
- Can be simplified for exponential family distributions.
- Can be even more simplified for Gaussian distributions.

Classic EM

- Auxiliary random variables

$$\tilde{\mathbf{x}}_i = \mathbf{s}_i(\boldsymbol{\theta}_i) + \tilde{\mathbf{n}}_i$$

- Noise

$$\mathbf{n} = \sum_{i=1}^{K} \tilde{\mathbf{n}}_i, \quad E\{\tilde{\mathbf{n}}_i \tilde{\mathbf{n}}_j^H\} = \beta_i \delta_{ij} \mathbf{\Pi}, \quad \sum_{i=1}^{K} \beta_i = 1$$

- *E Step*:

$$\widehat{\tilde{\mathbf{x}}}_i = \mathbf{s}_i(\boldsymbol{\theta}_i^k) + \beta_i(\mathbf{y} - \sum_{l=1}^{K} \mathbf{s}_l(\boldsymbol{\theta}_l^k))$$

- *M Step*:

$$\boldsymbol{\theta}_i^{k+1} = \boldsymbol{\theta}_i^k - (\boldsymbol{\nabla}_{\boldsymbol{\theta}_i} \boldsymbol{\nabla}_{\boldsymbol{\theta}_i}^T \phi_i(\boldsymbol{\theta}_i) + \lambda \mathbf{H}_i)^{-1} \boldsymbol{\nabla}_{\boldsymbol{\theta}_i} \phi_i(\boldsymbol{\theta}_i)|_{\boldsymbol{\theta}_i^k}$$

SAGE

SAGE: Space Alternating Generalized Expectation Maximization [Fessler and Hero, 94]

- Auxiliary random variable

$$\mathbf{x}^S = \mathbf{s}_i(\boldsymbol{\theta}_i) + \mathbf{n}$$

- *E Step*:

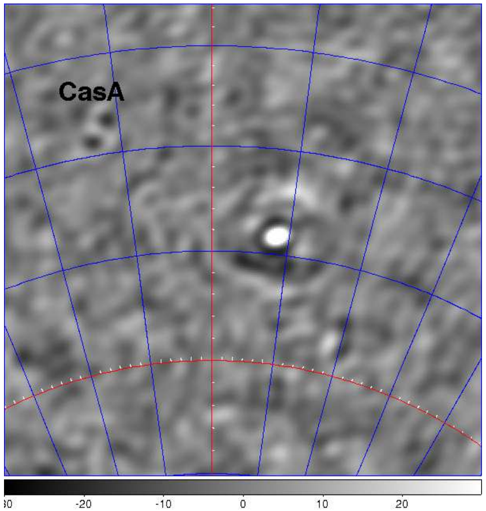
$$\widehat{\mathbf{x}^S} = \mathbf{s}_i(\boldsymbol{\theta}_i^k) + (\mathbf{y} - \sum_{l=1}^{K} \mathbf{s}_l(\boldsymbol{\theta}_l^k)) = \mathbf{y} - \sum_{l=1, l\neq i}^{K} \mathbf{s}_l(\boldsymbol{\theta}_l^k)$$

- *M Step*:

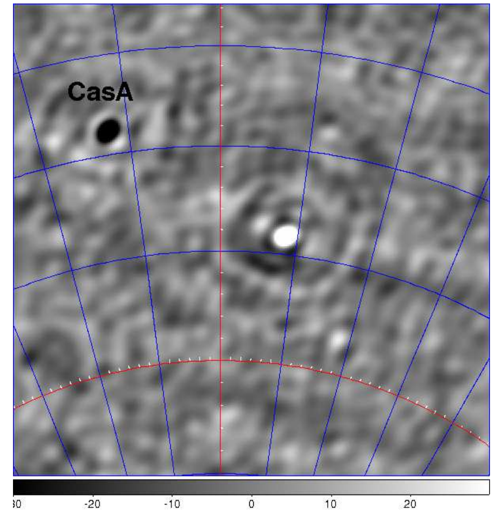
$$\boldsymbol{\theta}_i^{k+1} = \boldsymbol{\theta}_i^k - (\boldsymbol{\nabla}_{\boldsymbol{\theta}_i} \boldsymbol{\nabla}_{\boldsymbol{\theta}_i}^T \phi_i(\boldsymbol{\theta}_i) + \lambda \mathbf{H}_i)^{-1} \boldsymbol{\nabla}_{\boldsymbol{\theta}_i} \phi_i(\boldsymbol{\theta}_i)|_{\boldsymbol{\theta}_i^k}$$

- Caveat $f(\mathbf{y}, \mathbf{x}^S|\boldsymbol{\theta}) = f(\mathbf{y}|\mathbf{x}^S, \boldsymbol{\theta}_{\tilde{S}}) f(\mathbf{x}^S|\boldsymbol{\theta})$
- Faster convergence than the classic EM

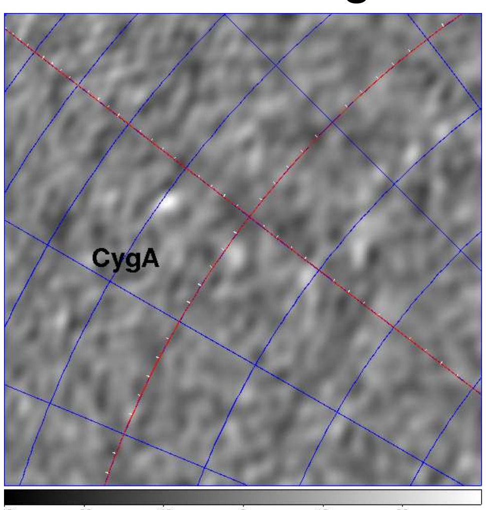
Example-1 (LOFAR)



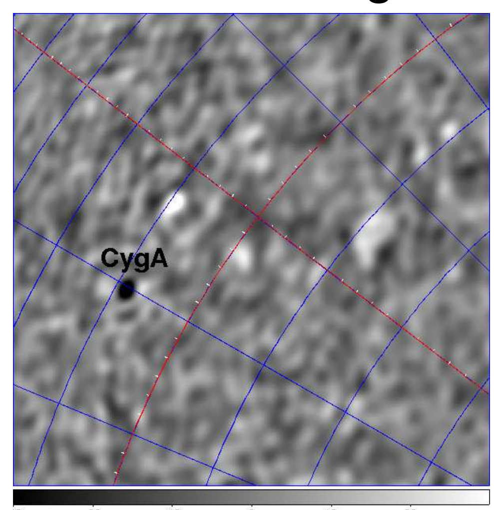
CasA, SAGE Algorithm



CasA, Normal Algorithm

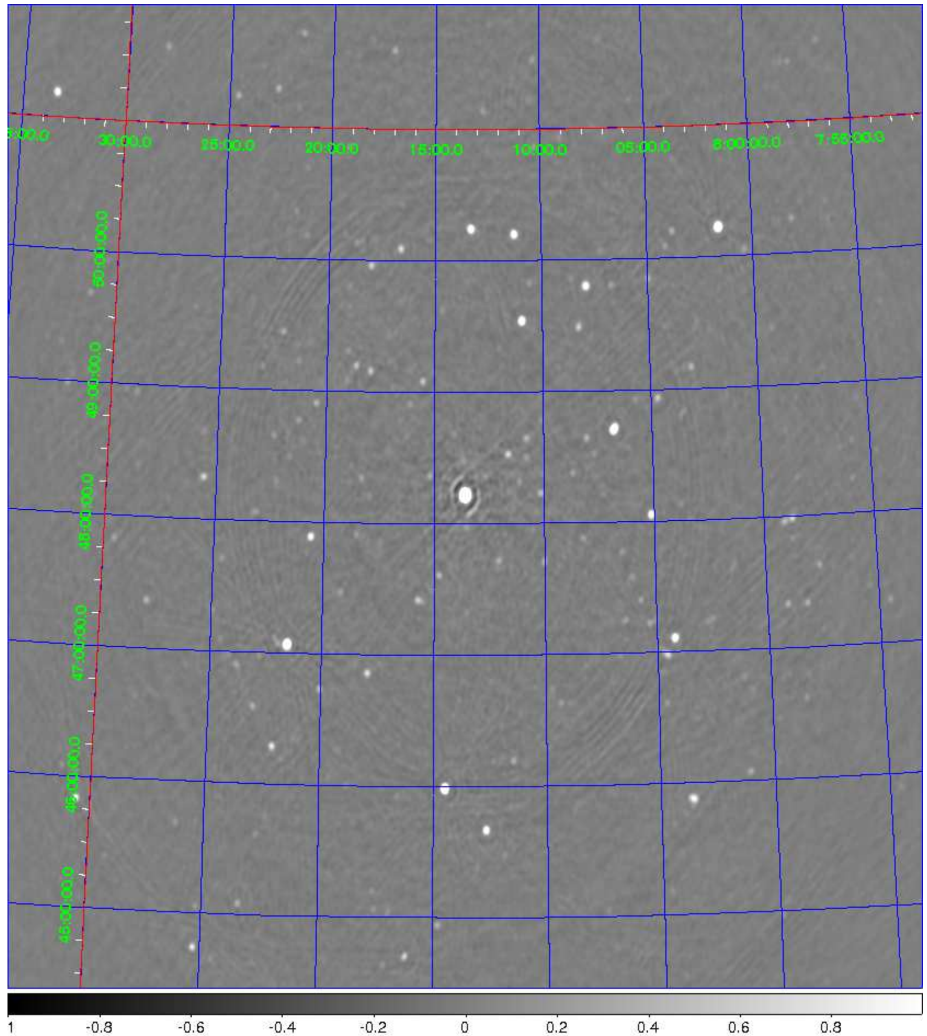


CygA, SAGE Algorithm



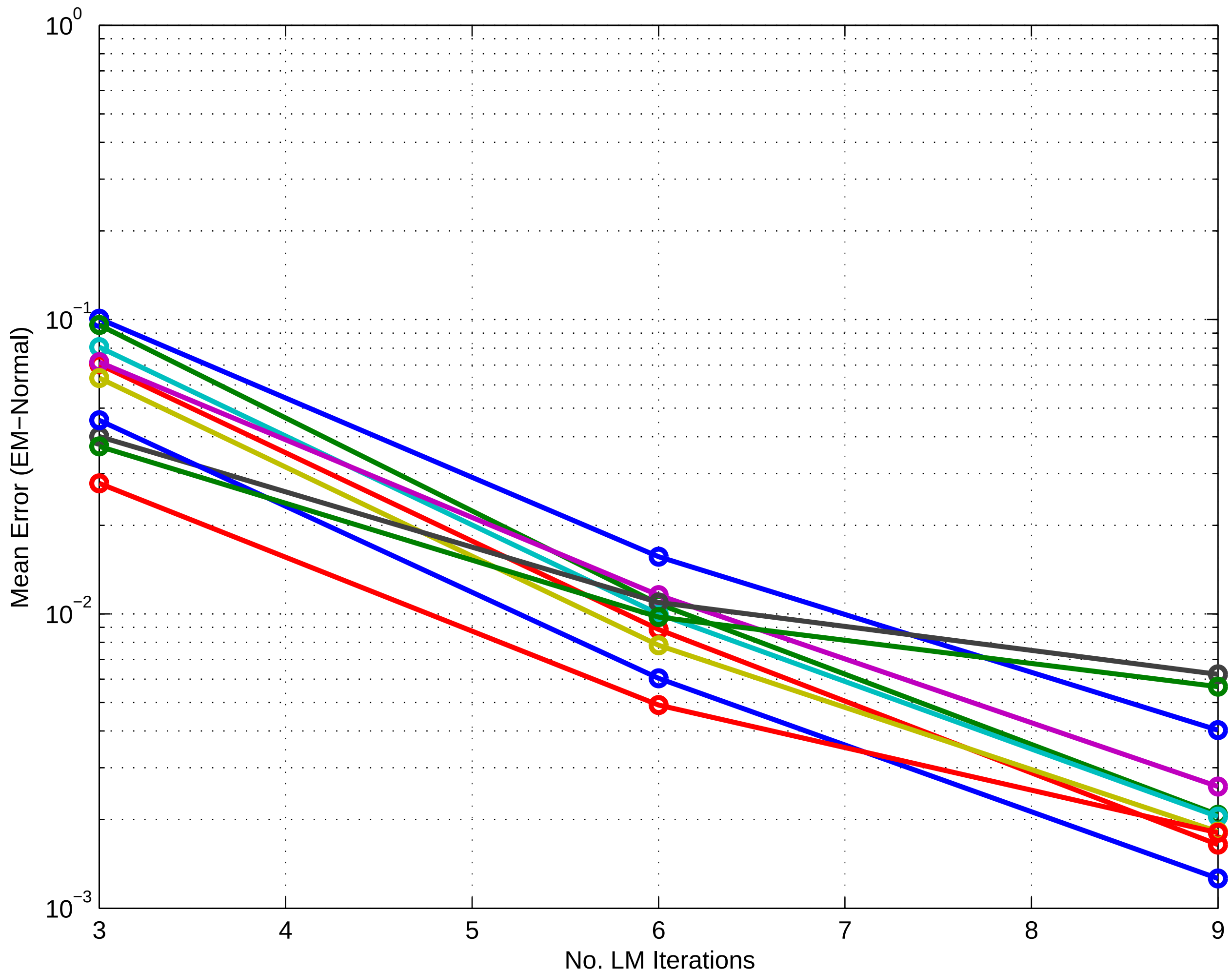
CygA, Normal Algorithm

Example-2



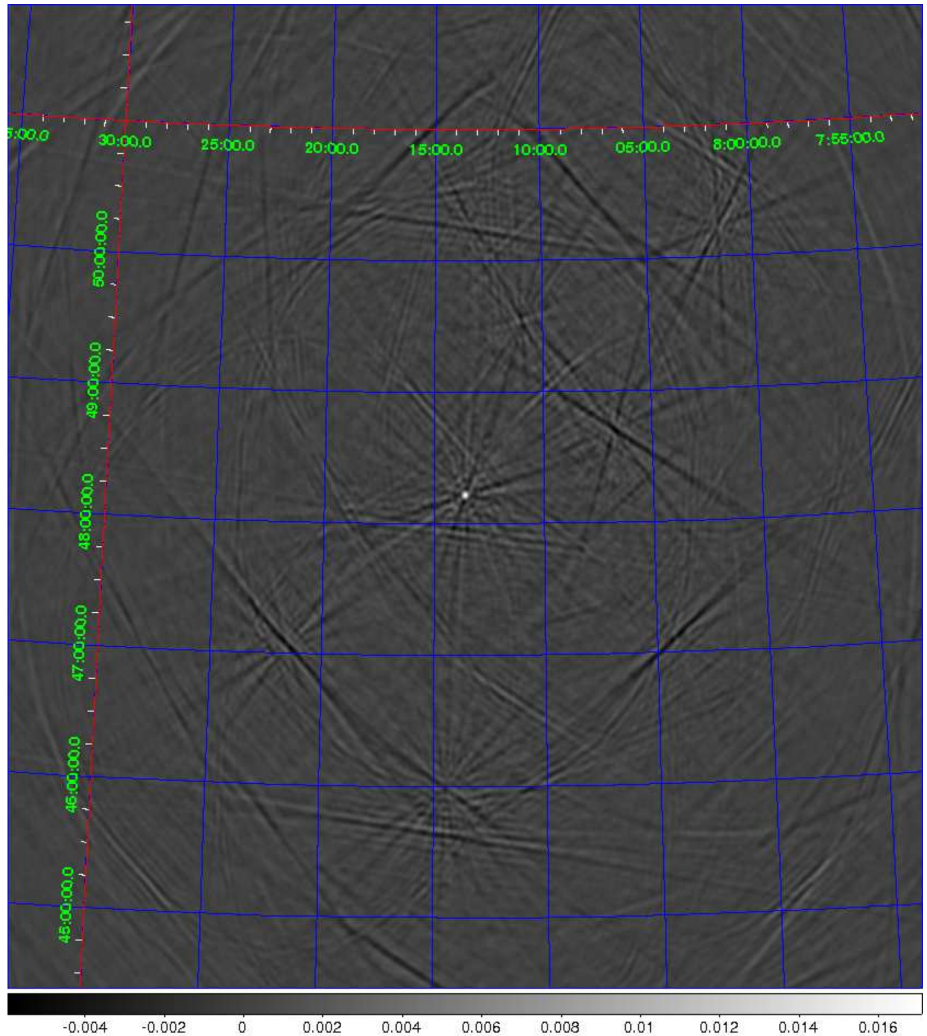
3C196 Field, WSRT

Example-2



Difference in Solutions

Example-2



Difference in Normal and SAGE algorithms

Conclusions

- EM/SAGE can reduce computational cost from $\mathbb{O}(K^2N^2)$ to $K\mathbb{O}(N^2)$ for KN parameters
- Can also improve quality of solutions when not converging
- See [Yatawatta,Zaroubi et al. 2008] for more information
- There is plenty more work to be done
- Advertisement: We are looking for a PhD student