Consensus optimization: The only way forward

Sarod Yatawatta

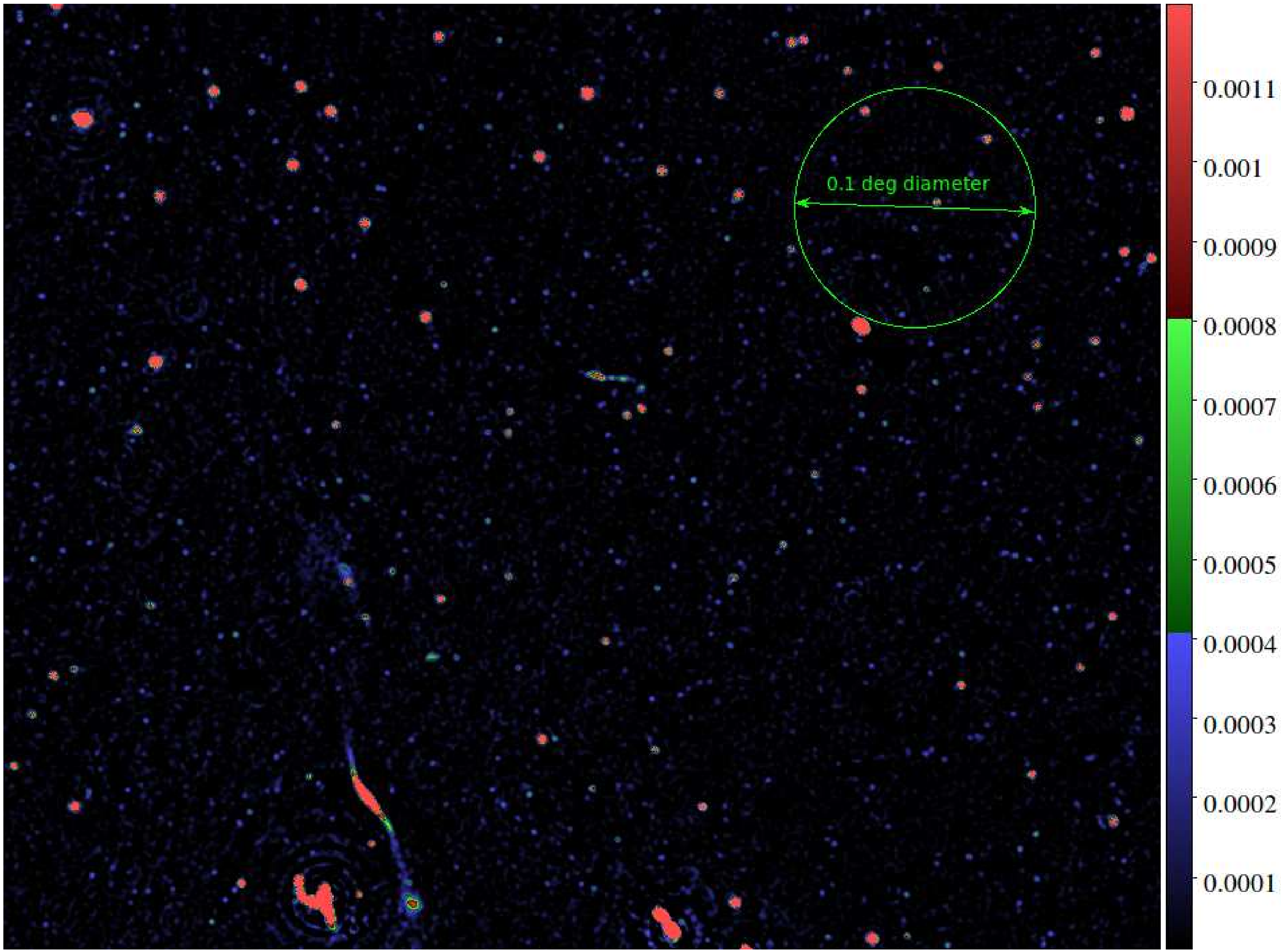
ASTRON

The Netherlands

Introduction

- Modern radio telescopes (LOFAR) produce large amounts of data.
- Calibration: essential for correcting systematic errors (beam,ionosphere), removal of foregrounds (Epoch of Reionization).
- Big data in radio astronomy: how can we calibrate large volumes of data, with minimal cost, while maximally exploiting smoothness in time, frequency and space?

LOFAR Deep Image



150 MHz, 2″ pixels, 40 μJy noise, dynamic range $>$ 150 000

Calibration

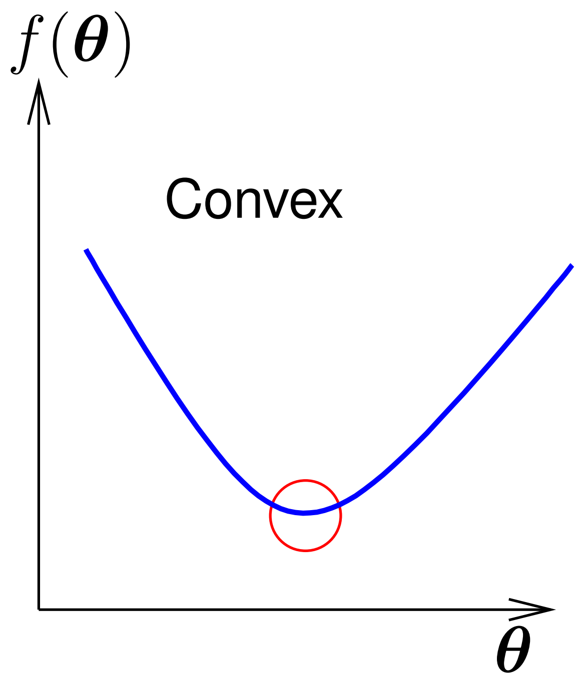
what we want after calibration

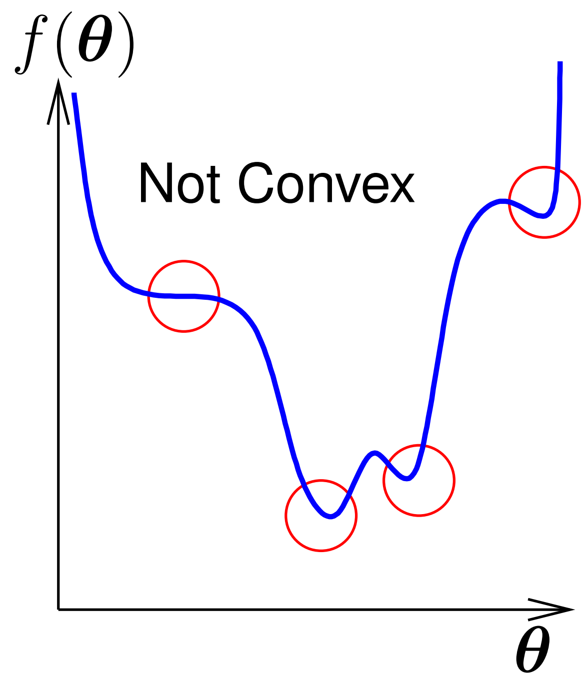
uncalibrated image

what we don't want

we certainly don't want this, though it's cute

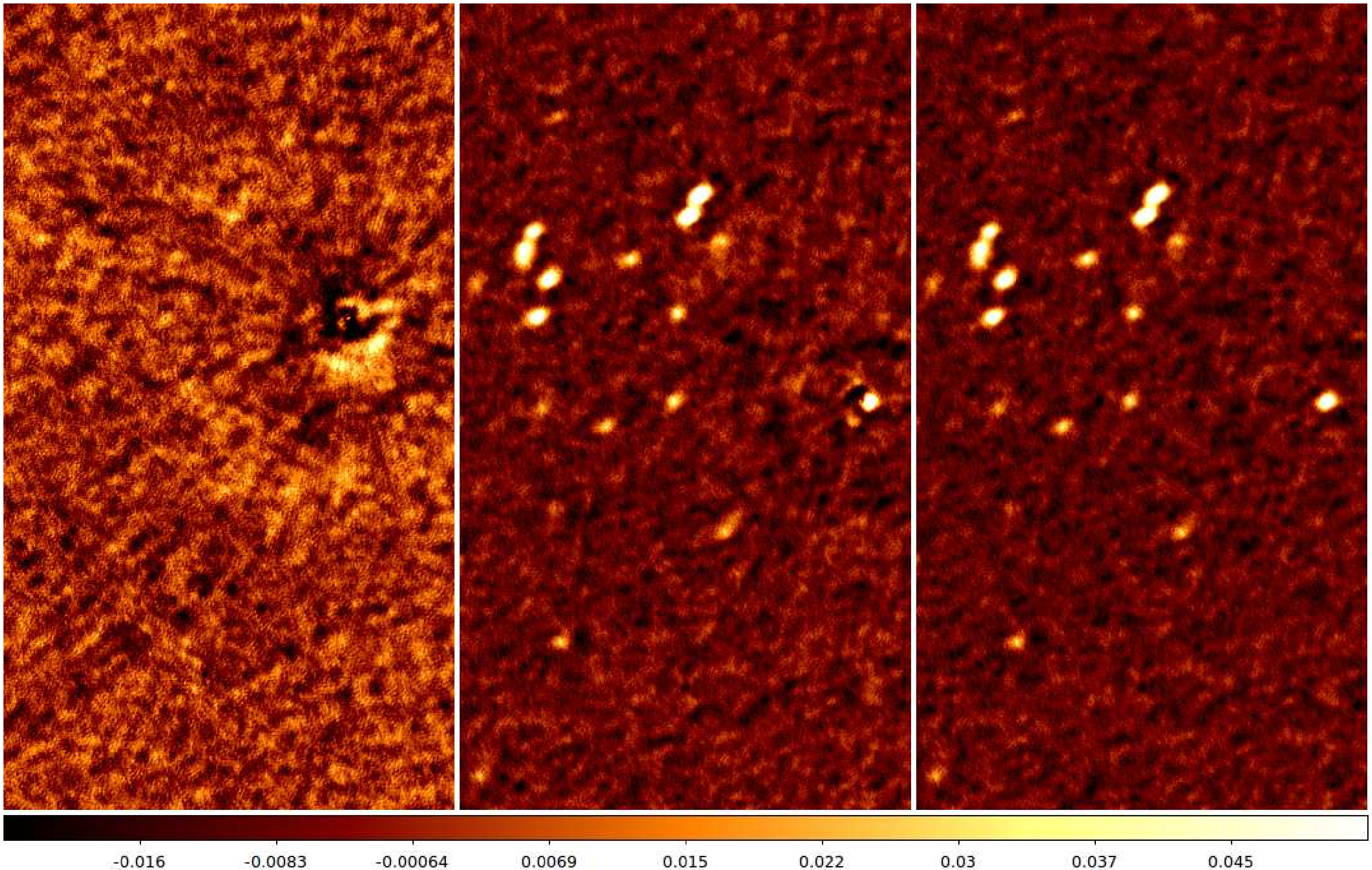
Calibration is Not Convex





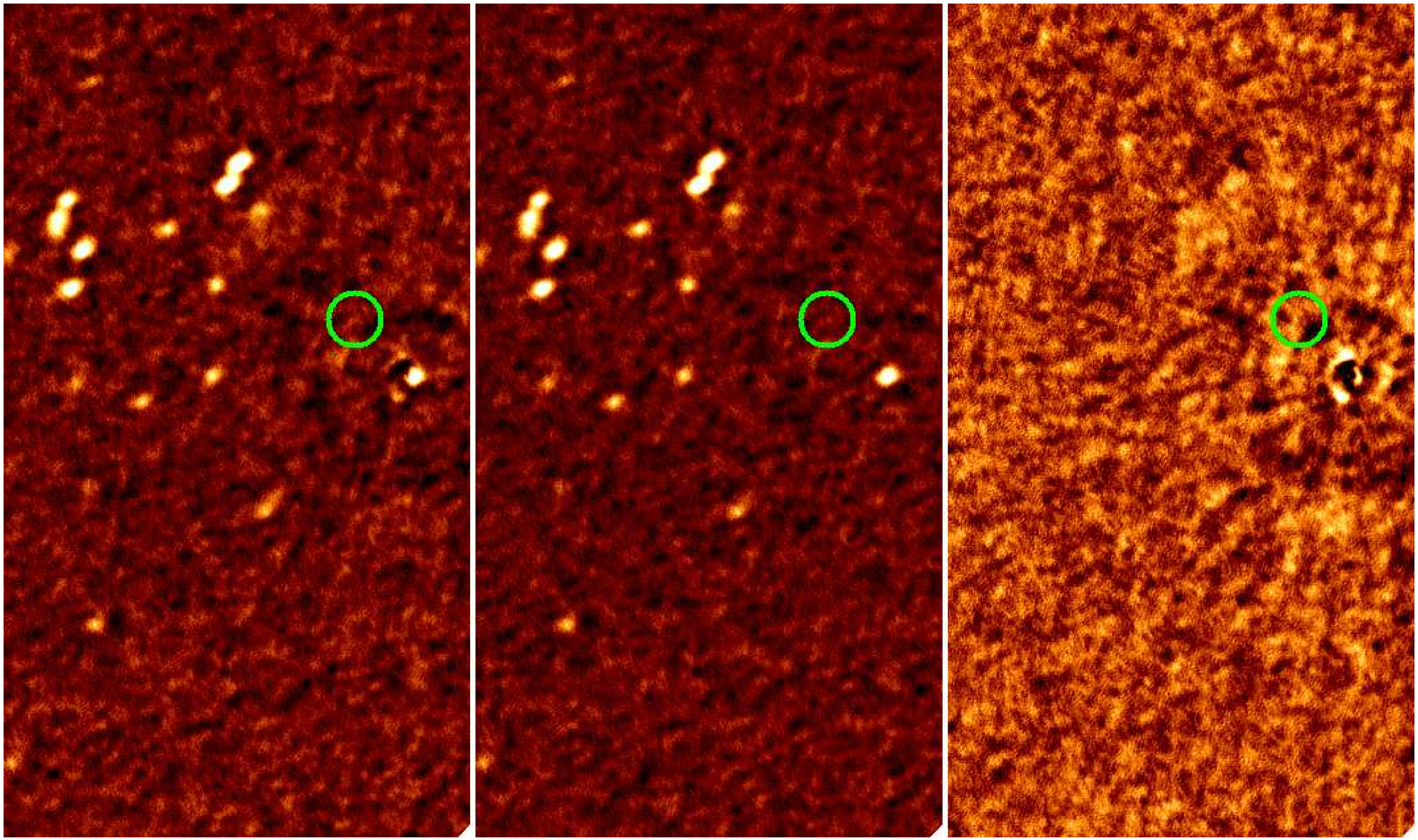
Solutions are on a 'compact' Riemannian manifold [Yatawatta, 2015]. A compact manifold does not have non-constant smooth convex functions [Udriste, 1994].
Example: phase-only calibration : solutions are on a hypersphere (compact) so not convex.

Simulated Example



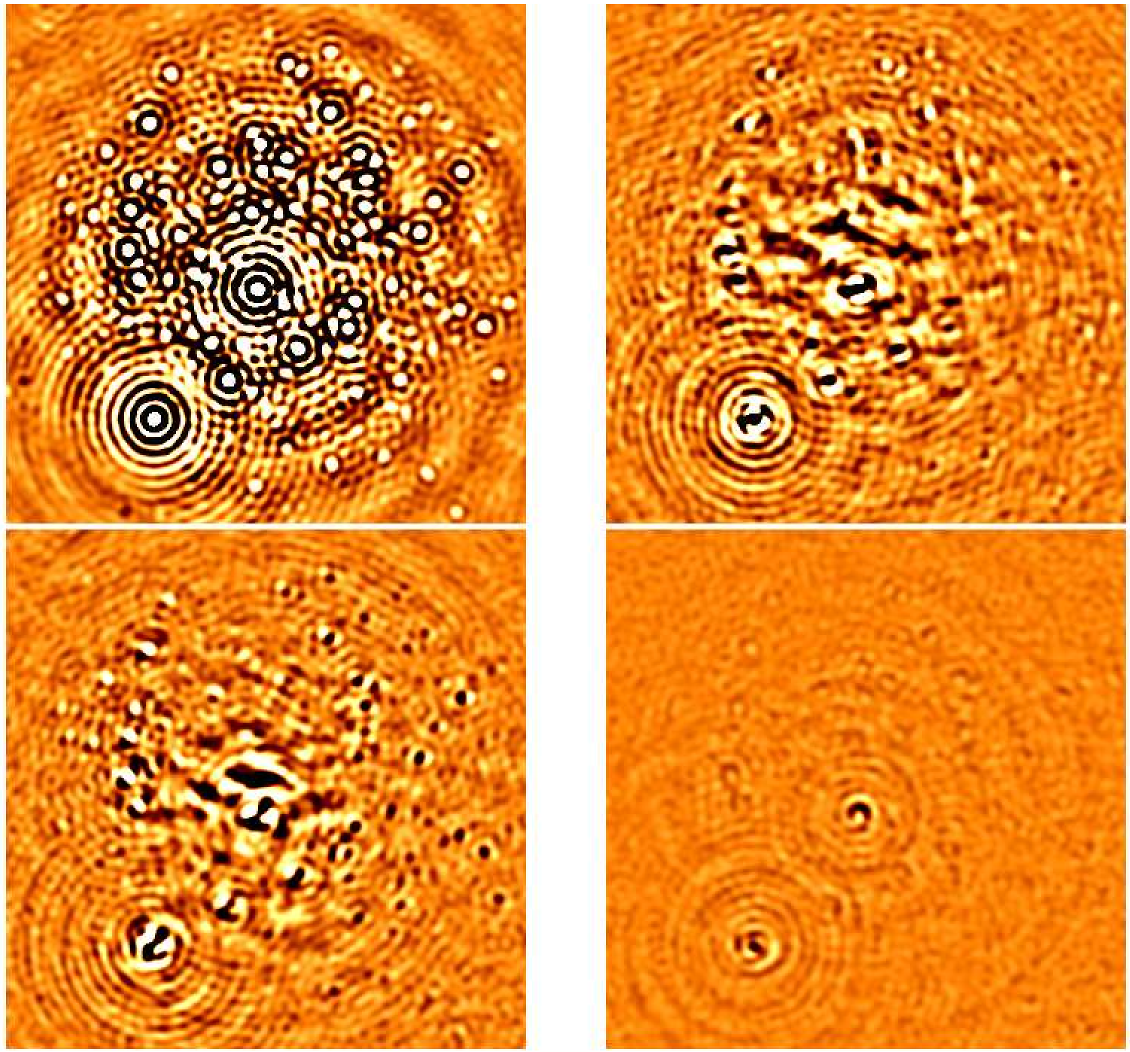
(left) raw data, (middle) local minima, (right) global minimum

Simulated Example



(left) local minima, (middle) global minimum, (right) difference=left-right
How can we increase the chances of getting the global minimum?

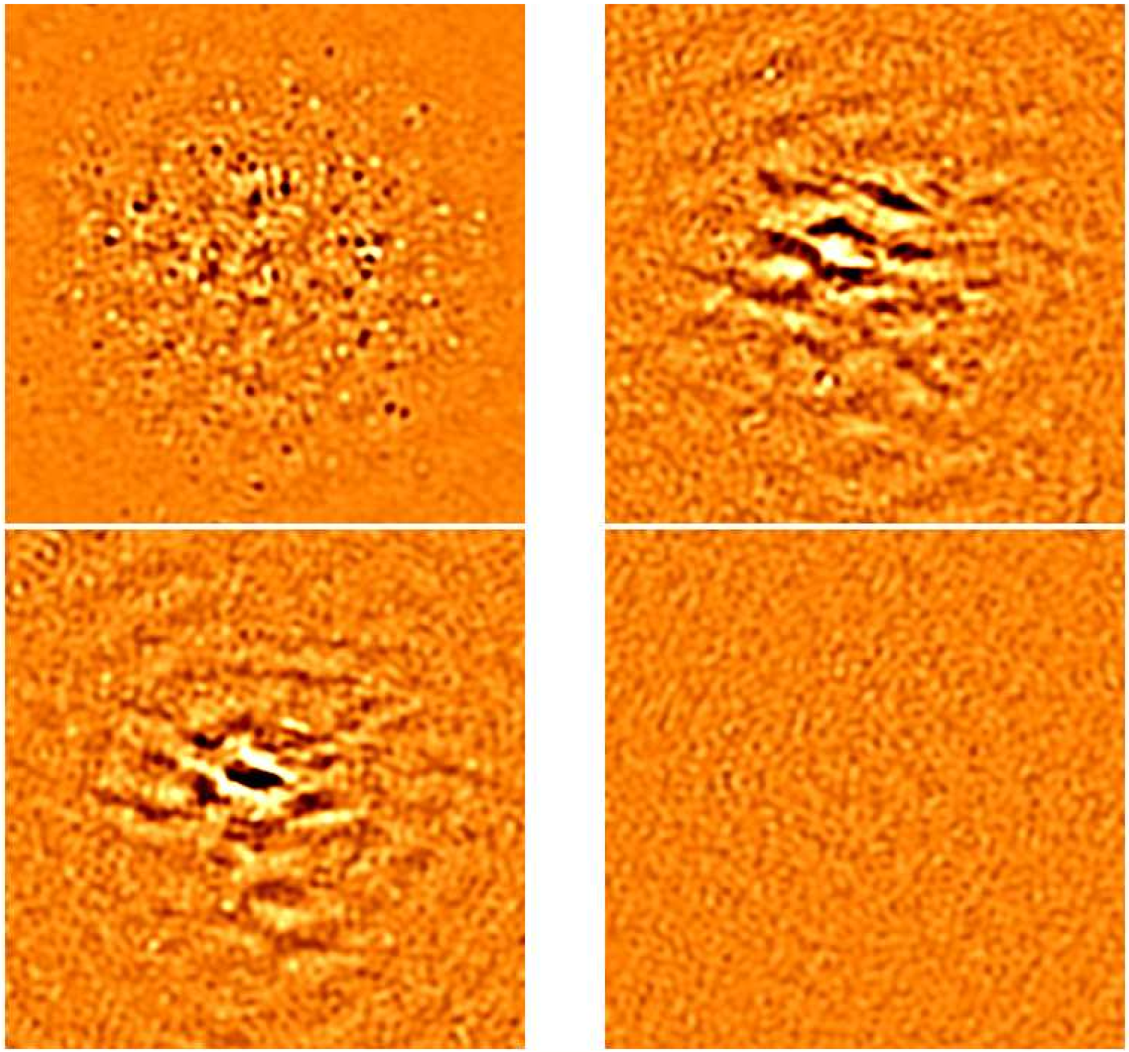
Image before calibration



I,Q,U,V images baselines $\leq$ 250 wavelengths

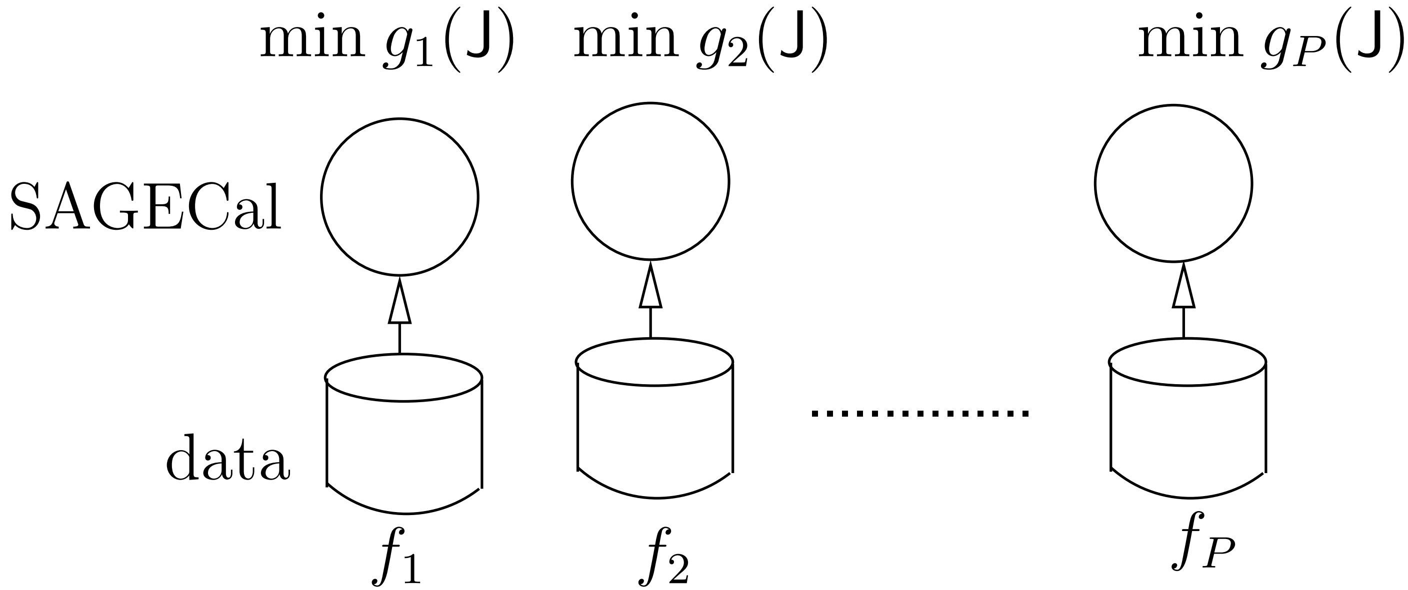
Sarod Yatawatta

Image after calibration



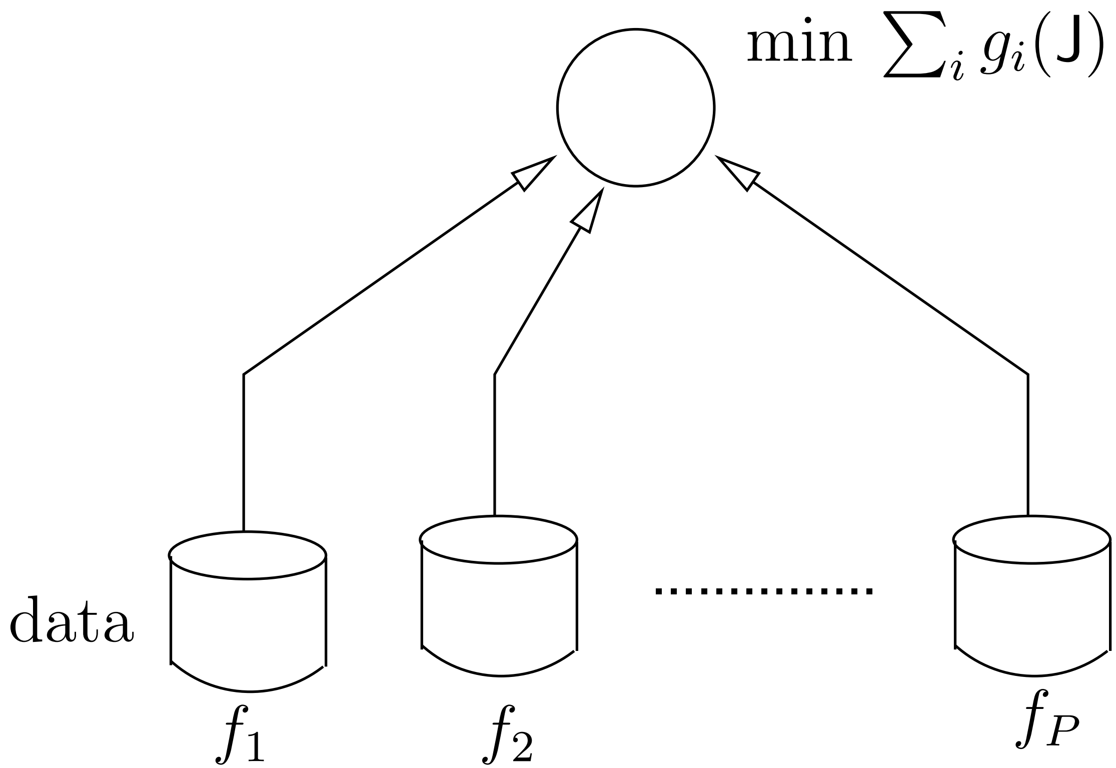
I,Q,U,V calibration using baselines > 250 wavelengths

Normal Calibration



Data distributed over a network of computers, divided into different subbands (frequencies).
Each SAGECal operates independently on data at different frequencies f_i. Solutions are only later interpolated.

What We Want



We want a unified solution exploiting smoothness in frequency,time and space (direction in the sky). This can avoid local minima as much as possible.
But this does not work in practice: too much data, not enough memory, no accurate model to parametrize.

Distributed Calibration

- Normal calibration: each SAGECal works independently

$$\mathsf{J}_{f_i} = \arg\min_{\mathsf{J}} \; g_{f_i}(\mathsf{J})$$

- Distributed calibration: each SAGECal appears to work independently, but actually solves

$$\{\mathsf{J}_{f_1}, \mathsf{J}_{f_2}, \ldots, \mathsf{Z}\} = \arg\min_{\mathsf{J}_{f_i},\ldots,\mathsf{Z}} \sum_i g_{f_i}(\mathsf{J}_{f_i})$$

$$\text{subject to } \; \mathsf{J}_{f_i} = \mathsf{B}_{f_i}\mathsf{Z}, \;\; i \in [1, P]$$

where J_{f_i} are local parameters, $\mathsf{B}_{f_i}\mathsf{Z}$ smoothing constraint across frequency,time and space.

- Basic principle is consensus optimization : details [Tsitsiklis,1984], [Boyd et al.,2011], [Yatawatta, 2015].

Consensus Optimization

Augmented Lagrangian

$$L(\mathbf{J}_{f_1},\dots,\mathbf{Z},\mathbf{Y}_{f_1},\dots)=\sum_i g_{f_i}(\mathbf{J}_{f_i})+\|\mathbf{Y}_{f_i}^H(\mathbf{J}_{f_i}-\mathbf{B}_{f_i}\mathbf{Z})\|+\frac{\rho}{2}\|\mathbf{J}_{f_i}-\mathbf{B}_{f_i}\mathbf{Z}\|^2$$

Iterative optimization with $n=1,2,\dots$

- ☐ Locally optimize to find

$$(\mathsf{J}_{f_i})^{n+1}=\arg\min_{\mathsf{J}} L_i\left(\mathsf{J},(\mathsf{Z})^n,(\mathsf{Y}_{f_i})^n\right)$$

- ☐ Globally find average (closed form solution)

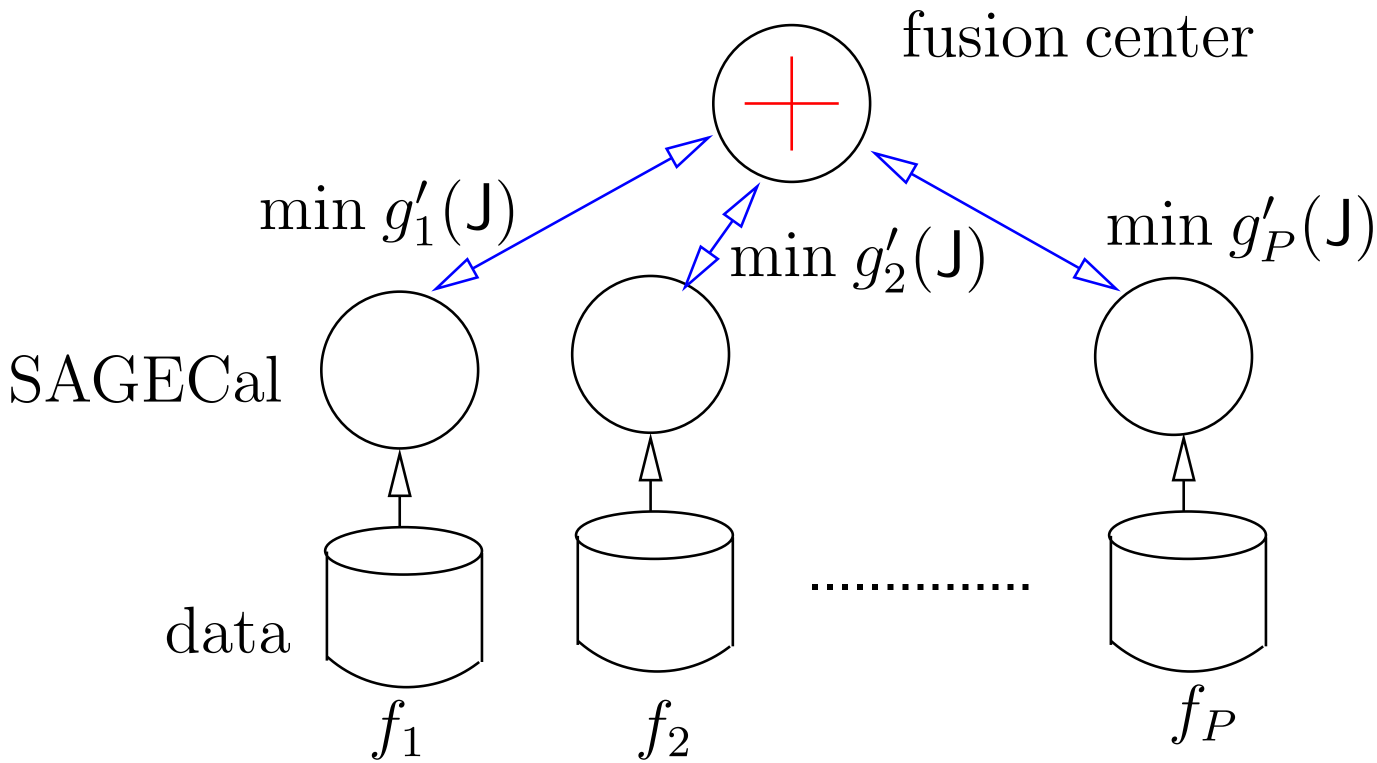
$$(\mathsf{Z})^{n+1}=\arg\min_{\mathsf{Z}}\sum_i L_i\left((\mathsf{J}_{f_i})^{n+1},\mathsf{Z},(\mathsf{Y}_{f_i})^n\right)$$

- ☐ Locally update Lagrange multiplier

$$(\mathsf{Y}_{f_i})^{n+1}=(\mathsf{Y}_{f_i})^n+\rho((\mathsf{J}_{f_i})^{n+1}-\mathsf{B}_{f_i}(\mathsf{Z})^{n+1})$$

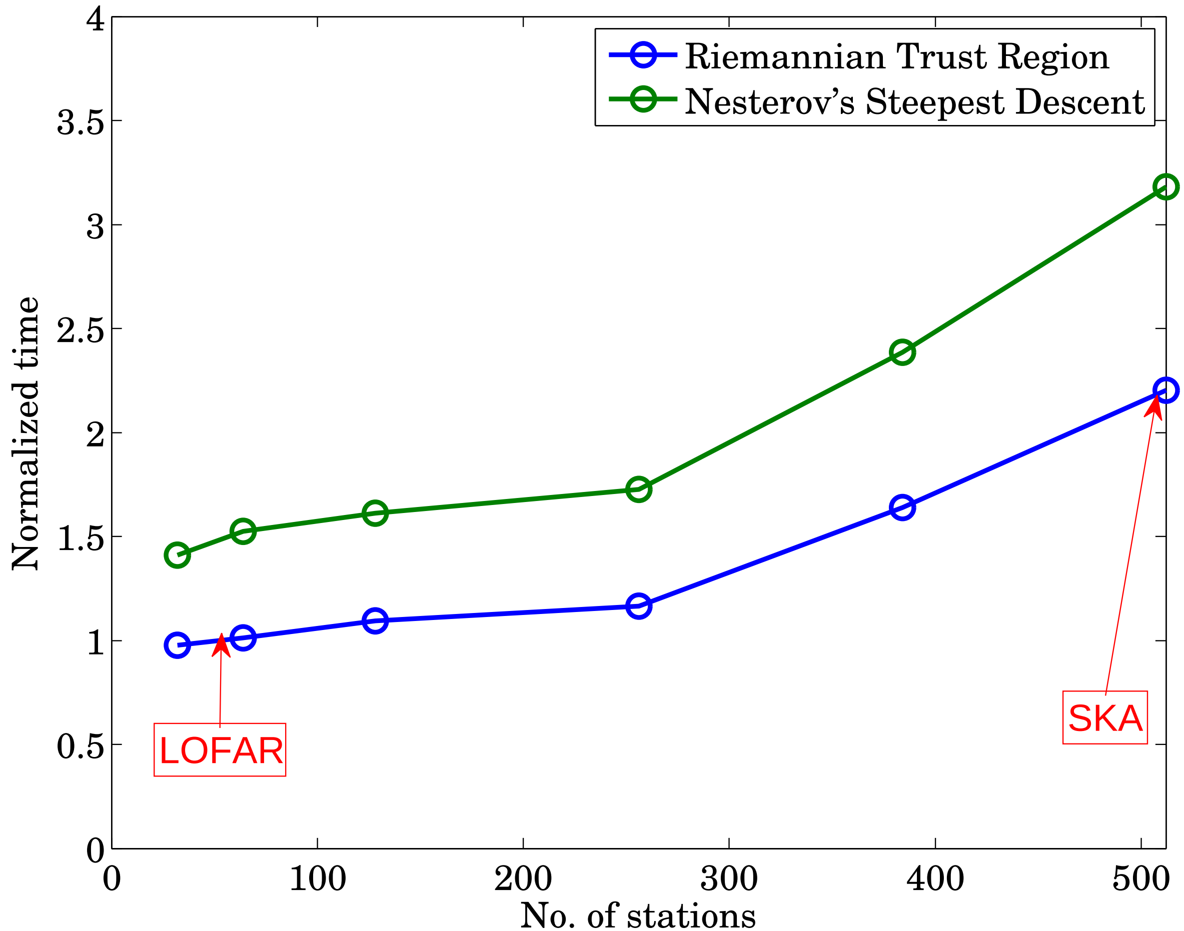
Large ρ makes the problem convex.

Distributed Calibration



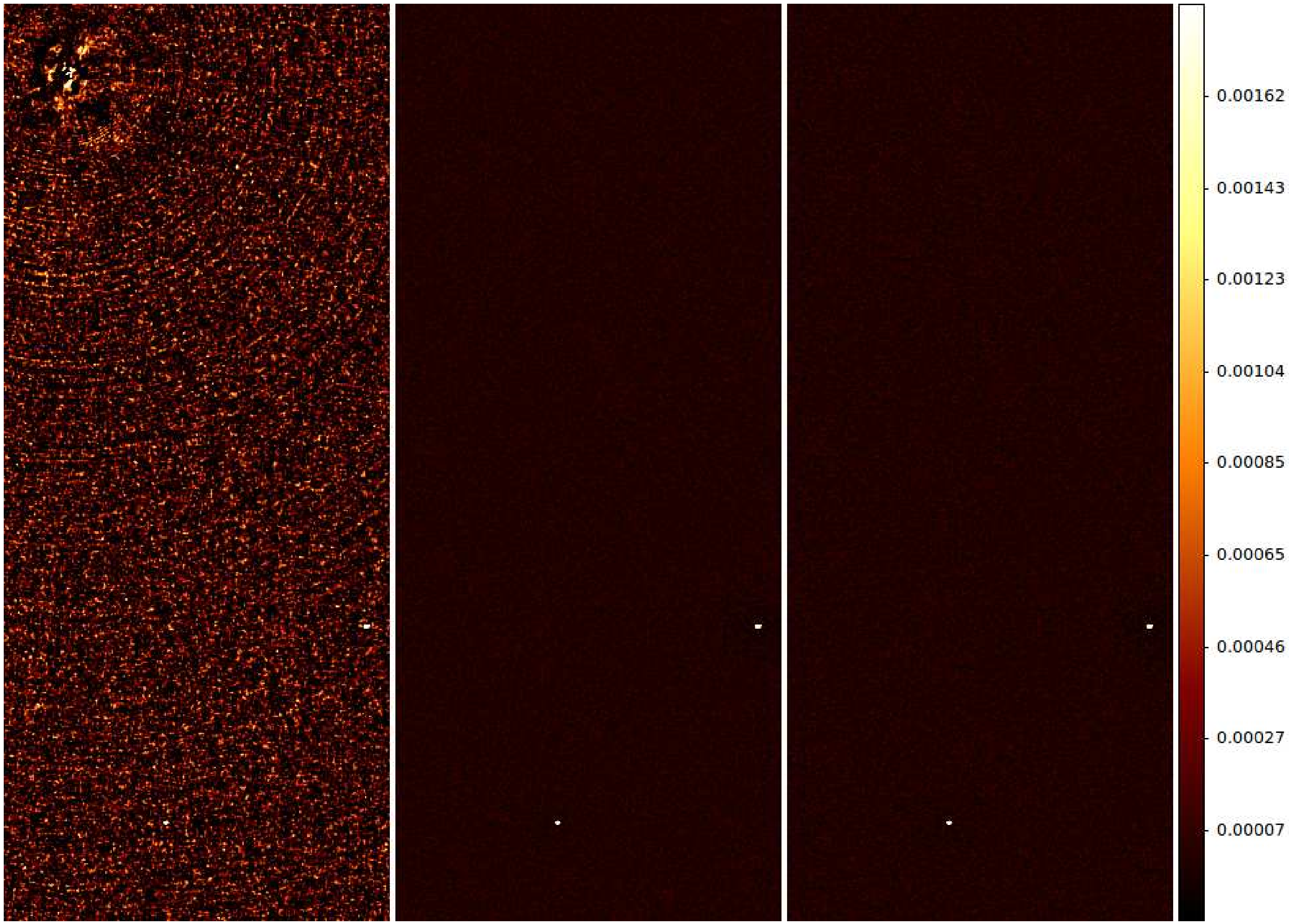
Information passed is much less than actual data calibrated (an order of magnitude less than other 'global' solvers).

Computational Time



Scaling from LOFAR to SKA-Low, 72 to 512 stations,
Linear scaling with the number of clusters (directions) calibrated.

512 Stations



(left) raw data (middle) RTR (right) Nesterov's

Conclusions

- ☐ Consensus optimization: can make calibration convex, thereby improving robustness.
- ☐ Can exploit smoothness in frequency, time and space with minimal computational cost and network communication cost.
- ☐ Almost linear scaling with number of directions calibrated and number of stations, almost constant cost with number of frequencies.
- ☐ Available at http://sagecal.sf.net .

Acknowledgments: European Research Council Advanced Grant LOFARCORE - 339743.