Aperture synthesis imaging

Theory of image reconstruction

Fabien Baron

University of Michigan

Monday 08 August 2012

PRINCIPLE

BAYESIAN APPROACH

The imaging problem

- Model fitting
  - Small number of components (often of different nature) needed to generate an image
  - Well-posed problem Ndata > Ncomponents
  - Constraints on parameters to keep them physical

- Image reconstruction
  - this is still model-fitting...
  - High number of identical components
    - e.g. pixels, wavelets, etc.
  - Ill-posed problem Ndata << Npixels
  - prior information needed to regularize the solution

Bayes theorem applied to imaging

- We want to find the most probable image given the data and given other assumptions we will call "imaging model"
- Bayes theorem for the *a posteriori* probability

prior

likelihood

$$\Pr(\boldsymbol{i}|\boldsymbol{D},\boldsymbol{M}) = \frac{\Pr(\boldsymbol{i}|\boldsymbol{M})\Pr(\boldsymbol{D}|\boldsymbol{i},\boldsymbol{M})}{\Pr(\boldsymbol{D}|\boldsymbol{M})}$$

marginal likelihood
const. for given reconst.

image

data

imaging model (image description, prior choices, ...)

Gaussian white noise on data $\Pr(\boldsymbol{D}|\boldsymbol{i},\boldsymbol{M}) \propto \exp\left[-\chi^2_{\boldsymbol{D}}(\boldsymbol{i})/2\right]$

Generic prior form $\Pr(\boldsymbol{i}|\boldsymbol{M}) \propto \exp\left[\mu R(\boldsymbol{i})\right]$

Regularized maximum likelihood

$$J(\boldsymbol{i}) = \chi^2(\boldsymbol{i}) + \mu R(\boldsymbol{i})$$

regularization weight

- If $\boldsymbol{i}$ is a collection of pixel fluxes $\{i_n\}$, the most probable image is the one that maximizes $J(\boldsymbol{i})$
- Additional constraints
  - positivity $\quad \forall n, i_n \geq 0$
  - normalization to unity $\quad \sum_n i_n = 1$

A reconstruction is defined by...

- Image model:
  - spatial basis (pixels, components, wavelets)
- Data:
  - Likelihood expression depends on noise model
  - Normal distribution in OIFITS standard
- Priors/regularization:
  - sets conditioning and remove degeneracies
  - based on user expectation in absence of data
- Minimization strategy
  - initial image
  - strategy/algorithm to iterate toward solution

LIKELIHOOD CHARACTERITICS

Likelihood Expression

- OIFITS assumptions
  - Noise on observables is normally distributed
    - Chi-squared statistics
  - Modulus and phases of bispectra are independent
  - Uncorrelated errors
- Power spectrum

$$\chi^2_{P_{ij}} = \frac{(|V_{ij}|^2 - P_{ij}^{\mathrm{data}})^2}{\sigma^2_{P_{ij}^{\mathrm{data}}}}$$

- Bispectrum

$$\chi^2_{B_{ijk}} = \frac{(|V_{ij}V_{jk}V_{ki}| - T_{ijk}^{\mathrm{data}})^2}{\sigma^2_{T_{ijk}^{\mathrm{data}}}} + \frac{(\arg\{V_{ij}V_{jk}V_{ki}\} - C_{ijk}^{\mathrm{data}})^2}{\sigma^2_{C_{ijk}^{\mathrm{data}}}}$$

Likelihood
Local minima

- In radio a likelihood made of complex visibilities has a single global minimum
  - due to the uniqueness of the phase
- In optical the atmosphere partially destroys the phase information, with N telescopes:
  - N(N-1)/2 phases but (N-1)(N-2)/2 closure phases
  - N-1 missing parameters per snapshot
    - (Lannes 2001, Meimon et al. 2009)
- In particular, an image translated by any amount still has the same squared visibilities and bispectra
  - the likelihood is multi-modal
- N>3, other effects than translation are involved...

Likelihood
A convex optimization problem ?

- Convex optimization regroups a wide variety of minimization algorithms

$$f(x^*) = \min\{f(x) : x \in \Lambda\}$$

$f(x) : \mathbb{R}^n \to \mathbb{R}$ should be a convex function

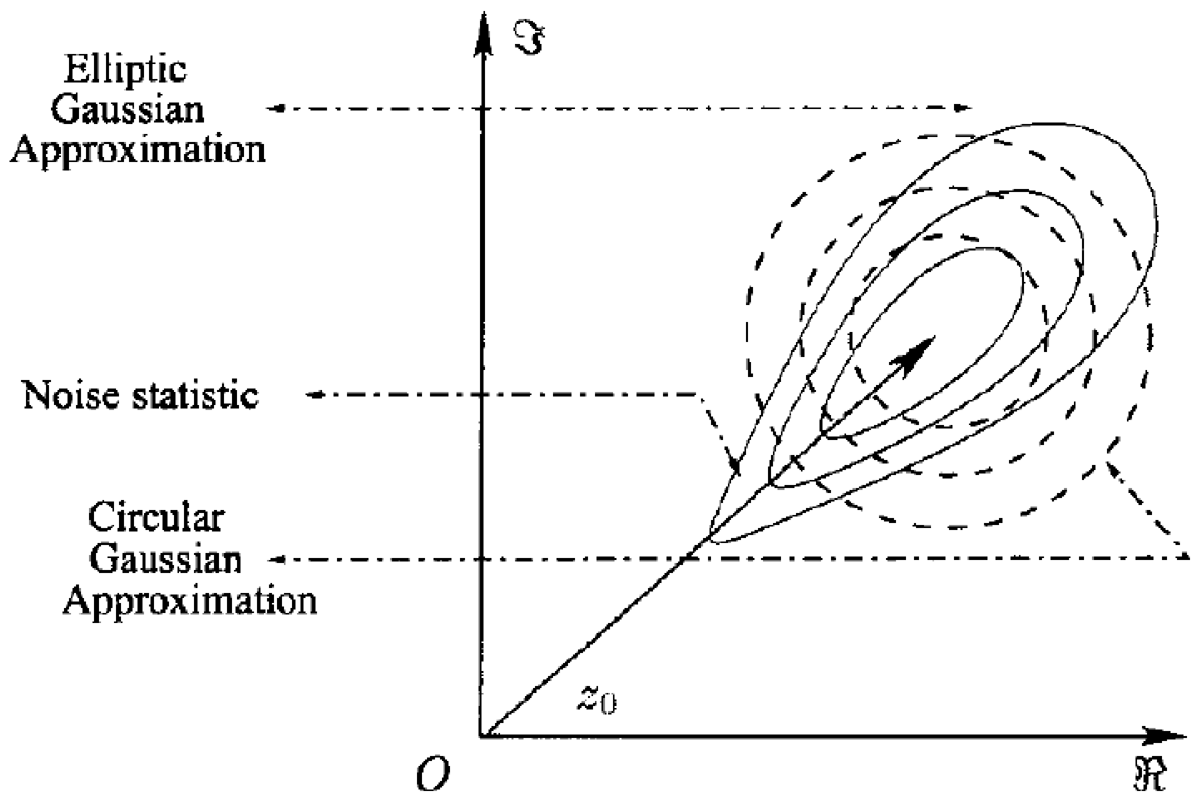
Λ should be a convex set

  - includes most gradient descent methods
  - found minima are global minima
- Convex set: a set of points containing all line segments between each pair of its points
- Convex function: a function whose graph forms a convex set
- Is the likelihood is convex ?

Likelihood
pears and bananas

- The "pear" bispectrum probability is convex

The circular approximation
is the Goodman model
Used in radio



Meimon et al., 2005

Likelihood pears and bananas

- The “banana” bispectrum probability is non-convex

Elliptic Gaussian Approximation required for convexity

Meimon et al., 2005

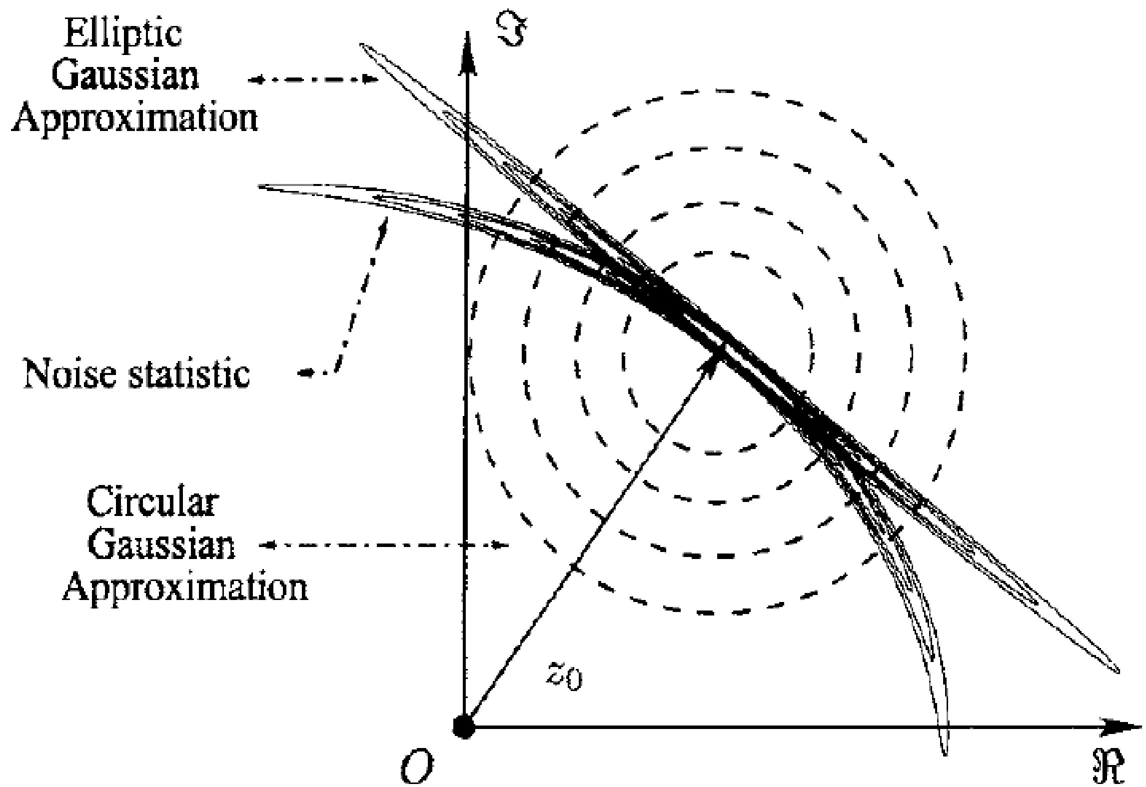
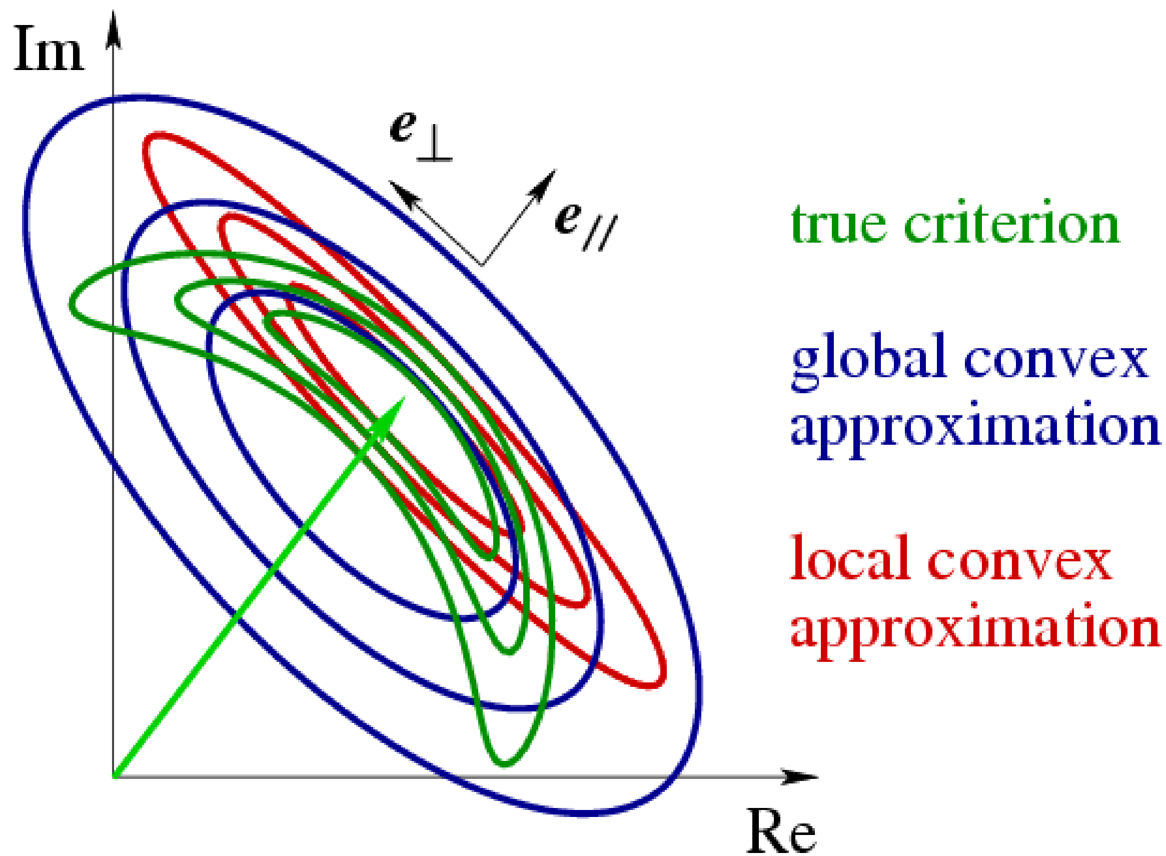


Fig. 2. Noise distribution contour lines, for $\sigma_r/|z_0| < \sigma_\varphi$.

rotation of the bispectrum
Vector
+ approximation of the
real and imaginary parts
of the rotated vector



$$\chi^2(\boldsymbol{i}) = \sum_k \left[\frac{\left(B_k - B_k^{\text{data}}\right).\,\boldsymbol{e}_{\text{rad}}}{\sigma^2_{\text{rad},k}} + \frac{\left(B_k - B_k^{\text{data}}\right).\,\boldsymbol{e}_{\tan}}{\sigma^2_{\tan,k}} \right]$$

REGULARIZATION

Regularization
Flux regularizers

- In absence of data, the image $\boldsymbol{i}$ should default to $\boldsymbol{i^0}$
  - $\boldsymbol{i^0}$ is called a prior image
  - most often chosen null, but not always
- Support constraint
  - outside region of interest, $\boldsymbol{i^0} = 0$
- A flux regularizer imposes this constraint using:

$$R(i) = \ell_p(\boldsymbol{i})^k \quad \text{or} \quad R(i) = \ell_p(\boldsymbol{i} - \boldsymbol{i^0})^k$$

$$\ell_p(\boldsymbol{x}) = \left(\sum_n |x_n|^p \right)^{\frac{1}{p}}$$

Why priors ?

- Needed because of the ill-posed nature of the reconstruction problem
    - Impose physical constraints
        - e.g. stellar disc diameter known
    - Lift likelihood ambiguities
        - fix center of image
    - Ease the reconstruction
        - “override” non-convexity

Regularization
Flux regularizers

$$\ell_0(\boldsymbol{i}) = \mathrm{Card}\{j; x_j \neq 0\}$$

number of non-zero pixels
Low and high fluxes are allowed
Sharp and mostly noise-free image

$$\ell_1(\boldsymbol{i}) = \sum_n |i_n|$$

Constant (normalization + positivity)
Useless

$$\ell_2^2(\boldsymbol{i}) = \sum_n i_n^2$$

Fluxes are penalized quadratically
High fluxes will be rare, smooth image
But less dynamic range, artefacts

$$R(\boldsymbol{i}) = \sum_n w_n i_n^2$$

$w_n \propto \theta$, compactness
(Gull-Skilling 1998, Le Besnerais 2008)

$$R(\boldsymbol{i}) = \sum_n i_n - i_n^0 - i_n \log \frac{i_n}{i_n^0}$$

Maximum entropy regularizer
$\ell_1 \ell_2$ behavior
Linear for high fluxes
Quadratic for low fluxes
Good dynamic range
Imposes positivity automatically

Regularization
Total variation

- Total variation (Chen 1998) imposes sparsity on the local gradient

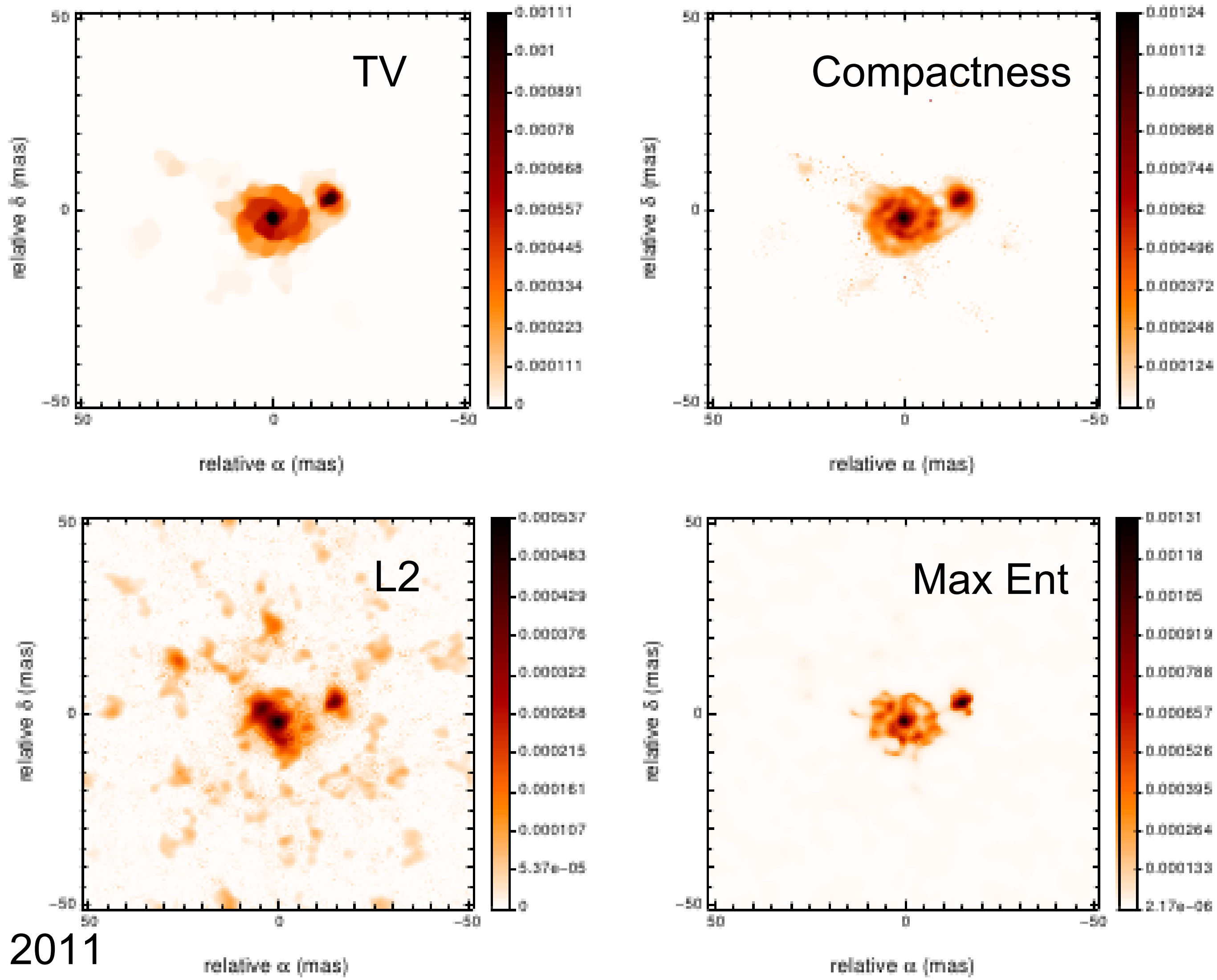
$$\ell_1(\boldsymbol{\nabla i}) = \sum_n |\nabla i_n|$$

- Favors patches of uniform fluxes separated by sharp transitions
- Local gradient is computed by the 5-point stencil method or simply by:

$$|\nabla i|^2 = (i_{n+1,m} - i_{n,m})^2 + (i_{n,m+1} - i_{n,m})^2$$

- This can be applied to reconstruction of mostly uniform objects
- The idea of sparsity will be presented in depth-tomorrow

Comparison of regularizers

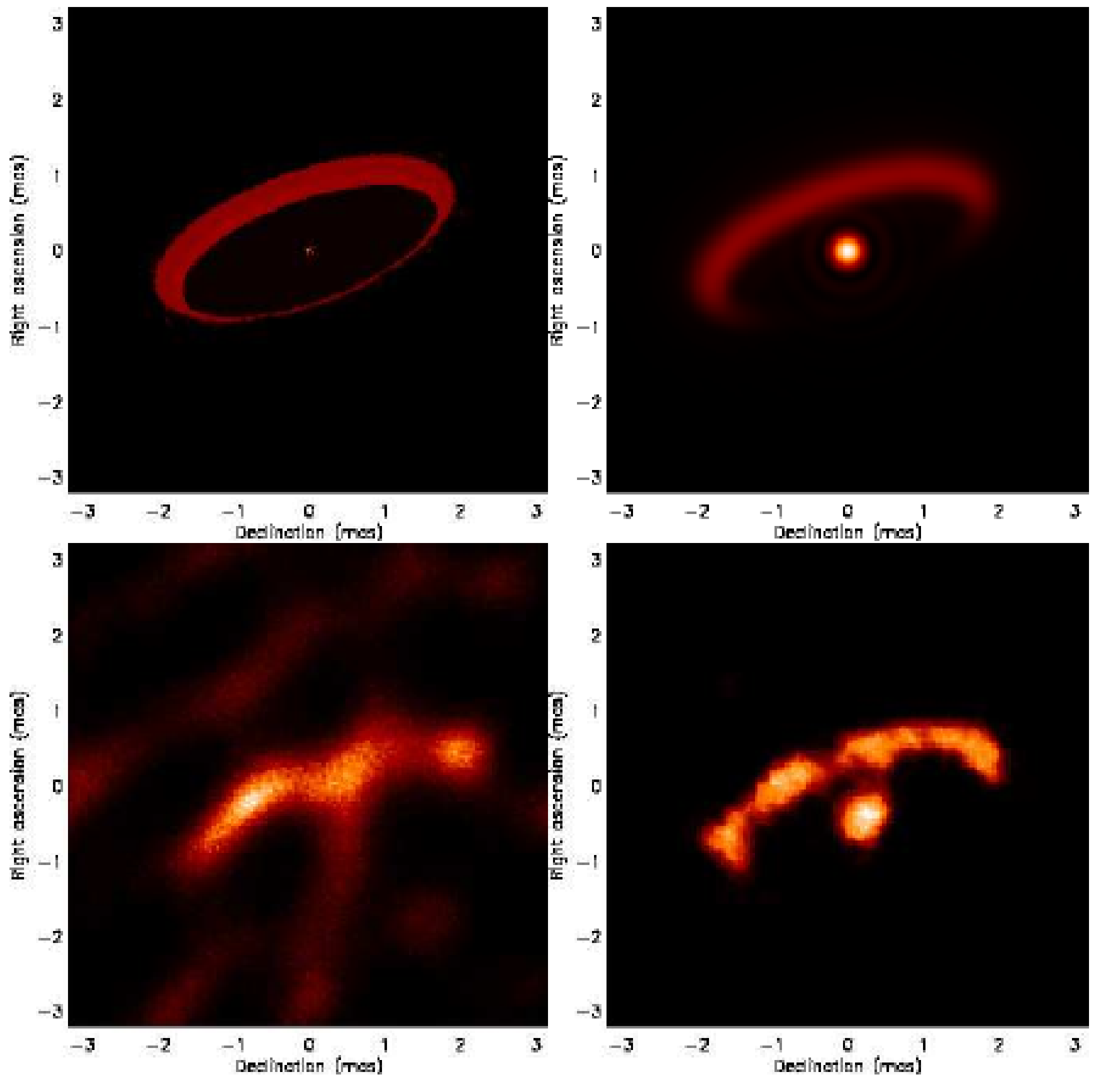


Renard et al. 2011

Comparison of regularizers

Truth

Convolved to array resolution

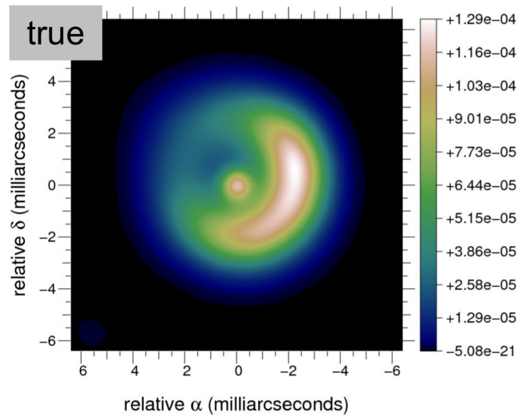


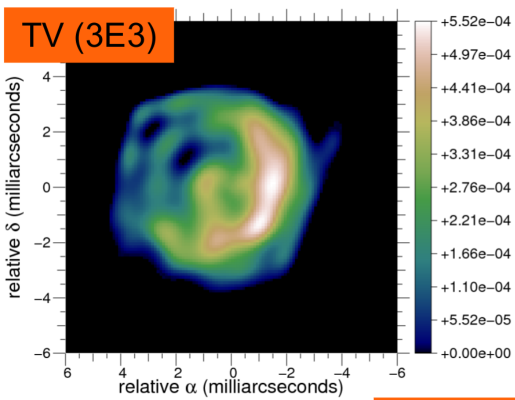
Maximum Entropy

Total variation

Regularization
Weight of regularization

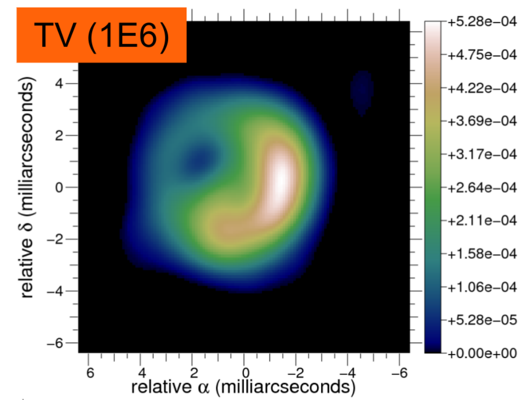
$$J(\boldsymbol{i}) = \chi^2(\boldsymbol{i}) + \mu R(\boldsymbol{i})$$

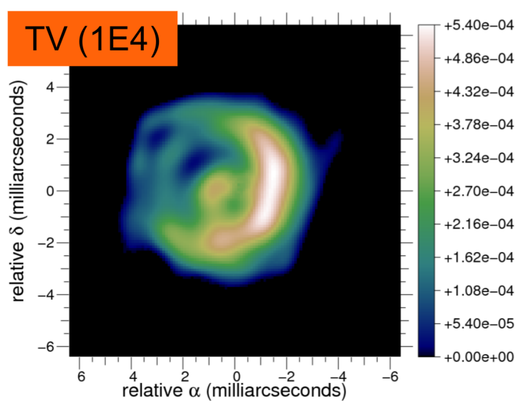


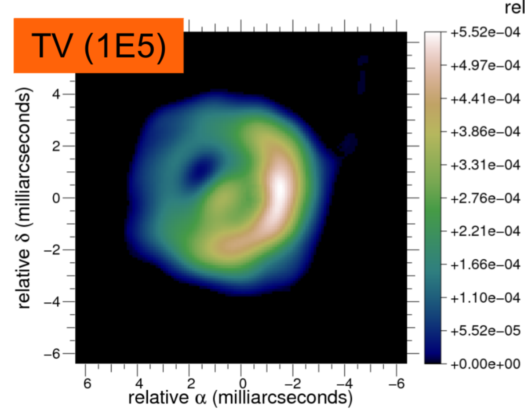


Rule of thumb

$$\chi_r^2 \simeq 1$$







Thiebaut 2008

MINIMIZATION METHODS

Minimization methods
Gradient descent

- Compute the gradient of criterion with respect to every image pixel
  - scale it by a reasonable factor
  - optionally, use an history of previous gradient
  - e.g.: steepest descent (BBM), conjugate gradients (IRS), semi-Newton with line search (MiRA), trust region method (BSMEM)
- Subtract this from the current image
  - but not for pixels that may become negative
- Repeat until convergence, based on:
  - variation of criterion from one image to the next
  - modulus of the criterion gradient
- Convexity is needed to converge to a global minimum...

Minimization methods
Monte Carlo Markov Chain

- Fluxes randomly move within the pixel grid
- Images with better criterion values are more likely to be accepted
- Simulated annealing/tempering: amount of movement is regulated by a temperature
  - Hot ? Fluxes move rather freely
  - Colder ? Image is settling progressively
- Supposed to find the global minimum
  - In practice, need good initialization

Aperture synthesis imaging

Main algorithms & packages

Fabien Baron

University of Michigan

Monday 08 August 2012

BBM/IRS

Building Block Method

- K.-H. Hofmann & G. Weigelt, 1993
- Image model : building blocks
  - One BB = delta function characterized by weight + location
- Data : bispectra

$$f_{\text{data}}(\boldsymbol{i}) = \sum_k w_k \left| V_{ab,k} V_{bc,k} V^{*}_{ac,k} - B_k^{\text{data}} \right|^2$$

$$w_k = \frac{\left| V_{ab,k} V_{bc,k} V^{*}_{ac,k} - B_k^{\text{data}} \right|^2}{\left| V_{ab,k} V_{bc,k} V^{*}_{ac,k} - B_k^{\text{data}} \right|^2 + Var\left(B_k^{\text{data}}\right)}$$

- Expression similar to Wiener filter
- Weights depend on searched parameters

Building Block Method

- Regularization :
  - sparsity, minimize the number of blocks (sharp images)

$$R_{\text{prior}}(\boldsymbol{i}) = \mu \, \|\boldsymbol{i}\|_0 = \mu \, \mathrm{Card}\left(\{ j ; i_j \neq 0 \}\right)$$

  - positivity and support constraint easily imposed
- Minimization : Matching pursuit
  - iterative reconstruction (linear approximation of the criterion, steepest descent on weight/location gradients)
  - basically CLEAN using bispectra...
- Negative blocks and multiple blocks can also be used
- Advantages:
  - Rather fast
  - Best software for sparse reconstructions
- Problems:
  - Most objects are not especially sparse...

BSMEM

BSMEM

- University of Cambridge (Buscher 94, Baron 2008)
- Image model : image pixels (+ 'experimental' wavelet coefficients)
- Likelihood : uses power spectra and bispectra
- Regularization: Maximum entropy
- Minimization method: Trust region method, local quadratic approximation of the criterion

$$J(\boldsymbol{i} + \delta\boldsymbol{i}) = J(\boldsymbol{i}) + \delta\boldsymbol{i}^T \nabla J + \frac{1}{2}\delta\boldsymbol{i}^T \nabla\nabla J \delta\boldsymbol{i}$$

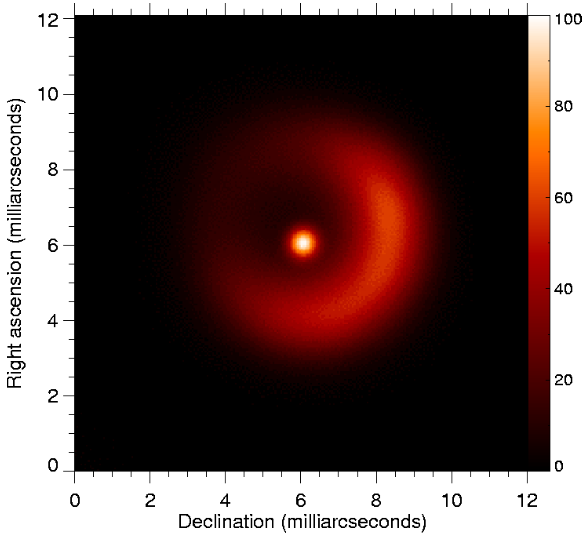
- Requires transposed transform and diagonal Hessian matrix
- Evaluation of the “trust region” in which this model is valid
  - Similar to Levenberg algorithm
- One step move to the minimum:

$$\delta\boldsymbol{i} = -(\nabla\nabla J + \beta\boldsymbol{I})^{-1}\nabla J$$

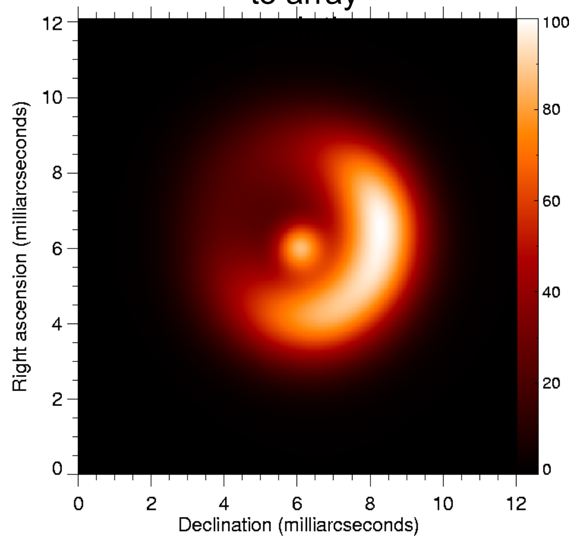
BSMEM

- Reconstruction of BC2004 data (Lawson et al. 2004)

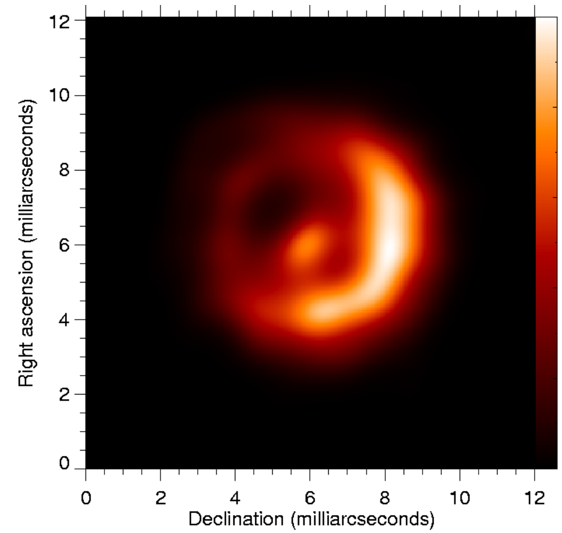
Truth image

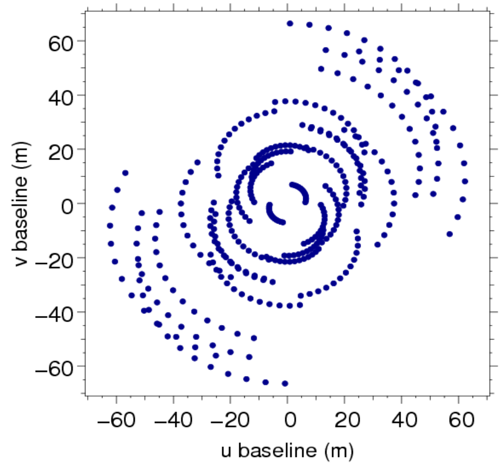


Truth convolved to array



BSMEM MAP image





BSMEM

- Advantages:
  - Speed
  - Reconstruction of binaries, because of fast exploration of the posterior
  - Automatic estimation of the regularization weight based on reduced chi-squared, or based on evidence (talk tomorrow)
- Problems:
  - Lack of regularization choices
  - Lack of control on the support of the reconstruction
    - Leading to artifacts...
  - Falls into local minima
  - Cannot do model-fitting simultaneously...

MIRA

MiRA

- University of Lyon (Eric Thiebaut, 2004)
- Image model : image pixels
- Likelihood : uses power spectra, bispectra, differential vis.
- Regularization: most common ones with gradients
- Minimization method: limited memory semi-Newton, with gradient projection onto positive definite domain
- Semi-Newton builds an approximation of the Hessian from the evolution of the gradient, using little memory
- Advantages:
  - Speed
  - Reconstruction of binaries, fast exploration of the posterior
  - **Vast** choice of priors
- Problems:
  - Falls into local minima
  - Cannot do model-fitting simultaneously...

MACIM

MACIM

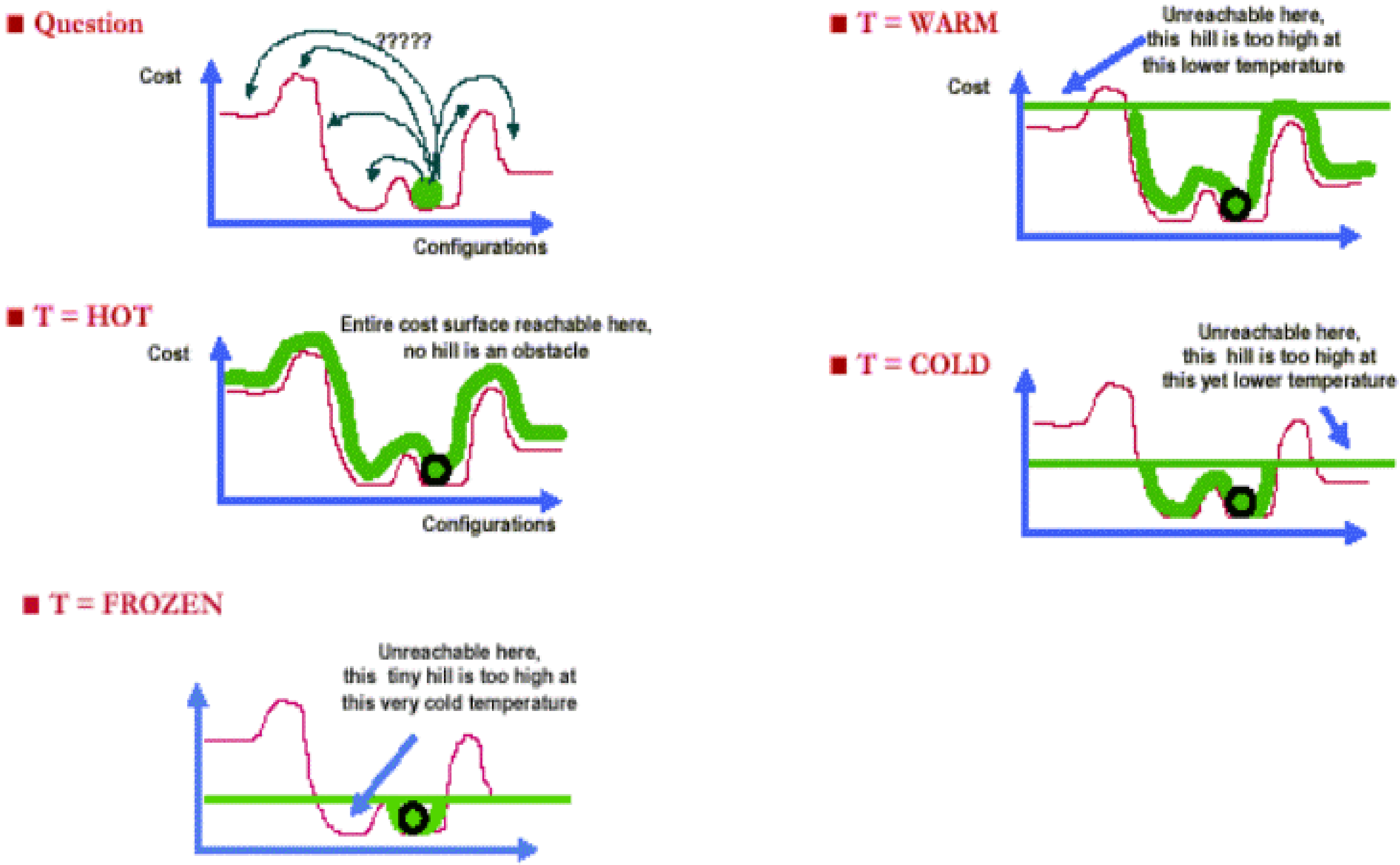
- MArkov Chain IMager, John Monnier +Mike Ireland (2006)
- Global optimization by stochastic exploration of the posterior probability
- Data : any, easily implemented
- Regularization : any regularizer, including those with no analytic gradients
- Minimization :
  - Markov Chain, simulated annealing = analogous physical process of heating and then slowly cooling a substance to obtain a strong crystalline structure (iron, chocolate...)

```
Repeat {
while NOT(minimum criterion at temperature T) {
      Perturb the image (add/move/swap/change pixels)
      Accept this new image with probability
```

$$p = \min\left[1, \exp\left(-\Delta\chi^2/2\mathrm{T} - \mu\,\Delta R_{\mathrm{prior}}\right)\right]$$

```
      } until criterion is minimum for T
      Decrease T according to cooling schedule
} until global minimum reached
```

MACIM – Simulated annealing



MACIM

- Advantages:
  - Simulated annealing search for the global minimum
  - Has limited model-fitting capabilities
    - joint estimation of any parameter possible in theory (ex: the regularization weight)
- Problems:
  - Choice of tuning parameters (cooling schedule, acceptance threshold)
  - Slower than local 'gradient based' optimization
  - Convergence to the global minimum in limited time not guaranteed
    - may still fall into local minima as the criterion is heavily multi-modal (newer samplers may solve this)
    - Initialization is important (e.g. hard to find both components of a binary)

WISARD

WISARD

- ONERA, Meimon et al. (2005 & 2009)
- Missing phase = visibility phase – closure information
  - if N stations, N-1 “differential pistons” parameters per snapshot
- Model : 'explicit' method = image pixels + missing phase parameters
- Data : pseudo-data of complex visibilities are formed
    - Visibility moduli + errors derived from powerspectra + errors
    - Phases + errors derived from closure phase data + missing phase parameters
  - likelihood based on rotated complex visibilities

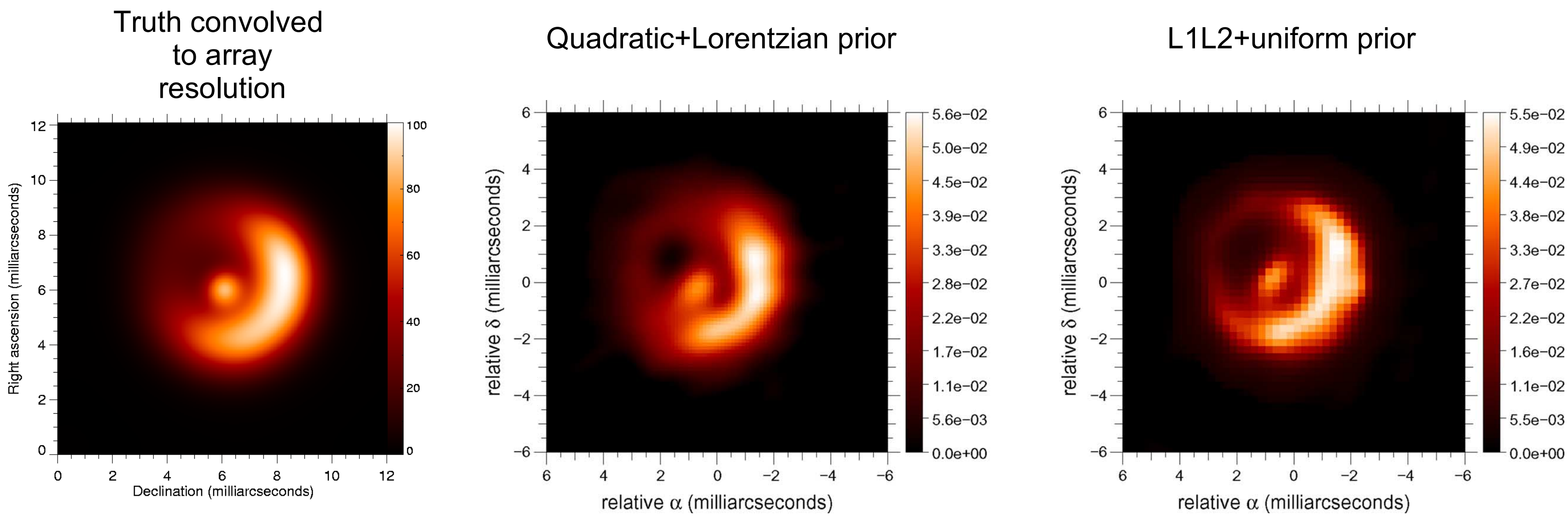
WISARD – Self calibration

- Alternating minimization scheme similar to self-calibration in radio

```
Step 1: initialization
Choose a starting image and regularization hyperparameter

Step 2: self-calibration
Find the optimal missing phase parameters for the current
image by minimising the likelihood (exhaustive search
possible for N=3)

Step 3: image reconstruction
Find the optimal image for the current missing phase
parameters by minimising the full criterion (likelihood +
regularization)

Repeat 2 and 3 till convergence
```

WISARD

- Regularization: same vast choice as MIRA



- Advantages
  - useful in weak phase case (N<5)
- Problems
  - Orphan powerspectra unused, bispectrum moduli unused
  - Quite slow
  - Comparison with MIRA shows very little difference
  - Self-calibration convergence depends on SNR...

Summary – Algorithmic strategies

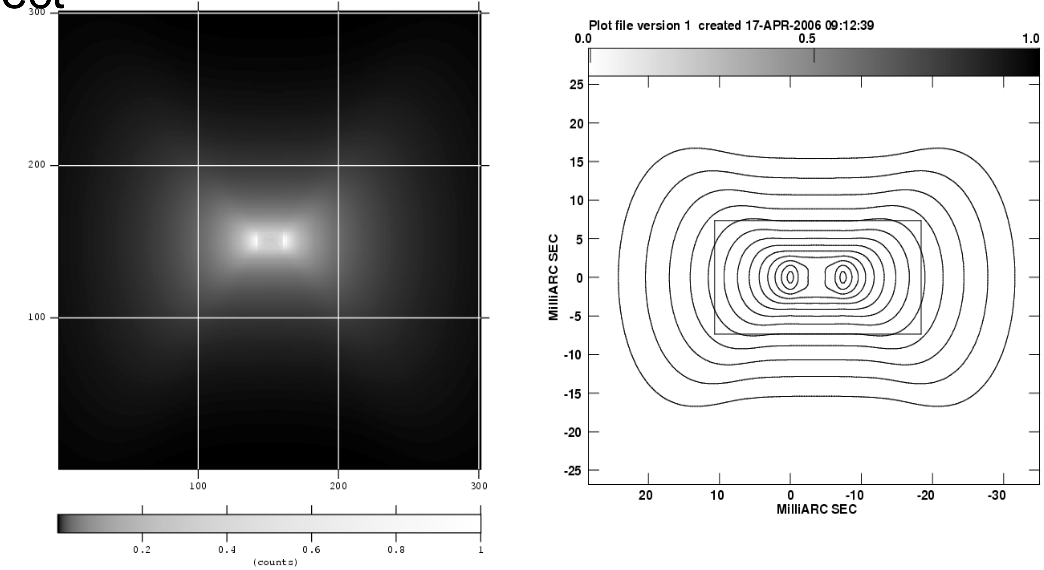
	MIRA	BSMEM	WISARD	BBM/IRS	MACIM
Data	All	Bispectra Powerspectra	Pseudo-complex visibility data formed from phase closures + powerspectra	Bispectra	All
Regularization	Positivity L1, TV, L2, L1L2, maximum entropy floating priors	Maximum entropy with prior image Wavelets	~ same as MIRA	Positivity Sparseness	Any type of regularization
Minimization scheme	Bound-constrained semi-Newton gradient descent	Trust region method (+unsupervised hyperparameter control)	Self calibration	Matched Pursuit (BBM) Conjugate gradient (IRS)	Simulated annealing

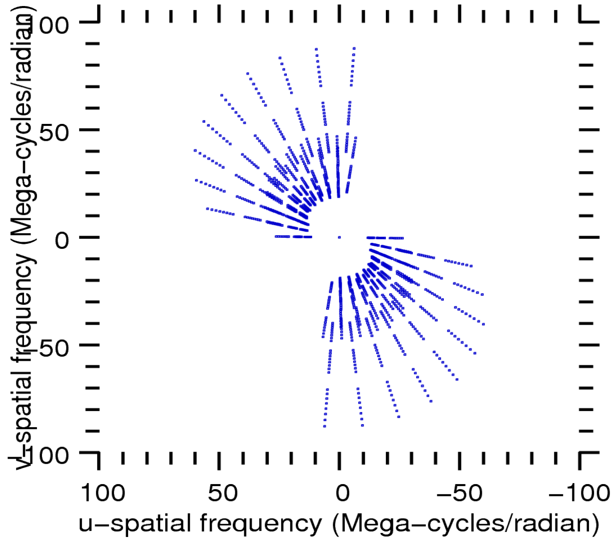
- Main difference = flexibility

Imaging Beauty Contest 2006

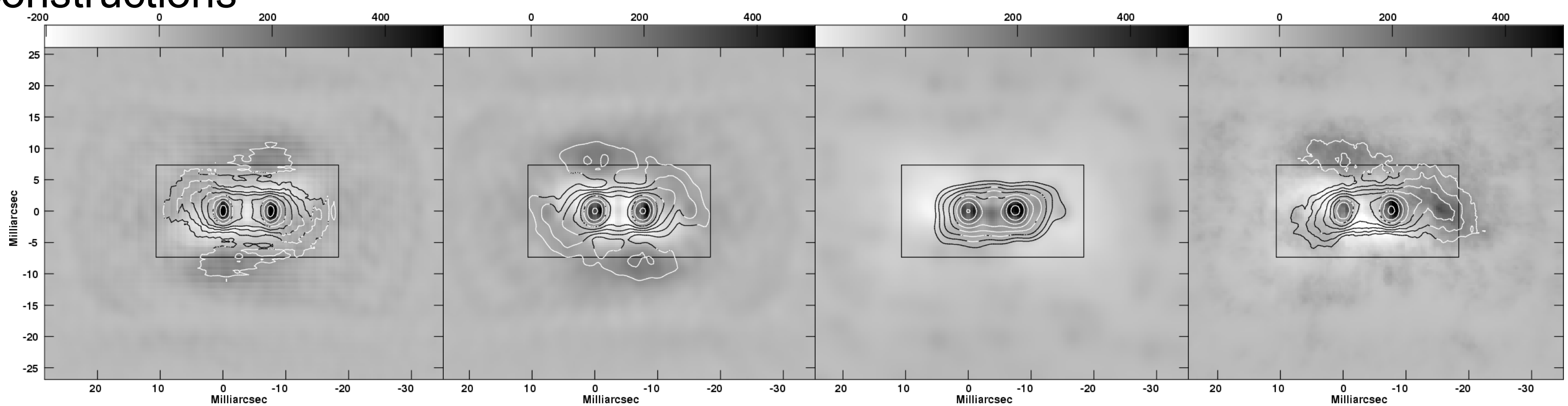
(Lawson *et al.*, 2006)

Object





Reconstructions

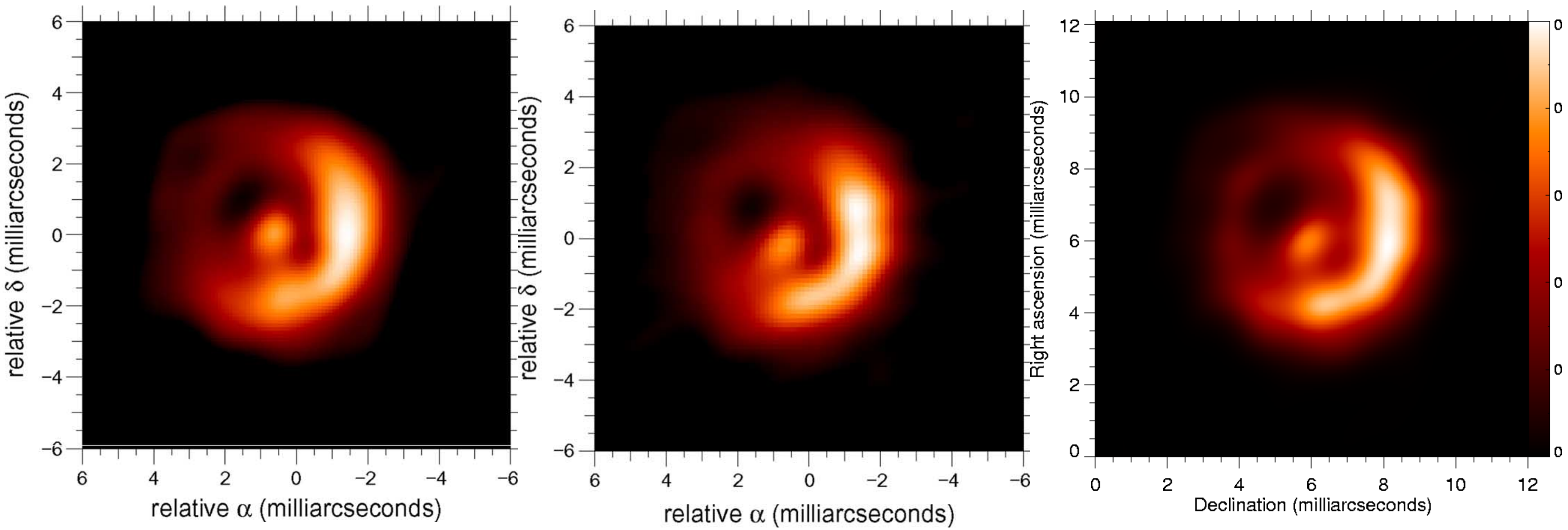


BSMEM MIRA BBM MACIM

Fabien Baron – Aperture synthesis imaging

MIRA, BSMEM, WIZARD on LkHa

Fabien Baron – Aperture synthesis imaging



- Giovanelli 2008, Baron 2008
- Feature comparison
  - same results with the same type of regularization
- More regularization/image model choices = more adaptability

Summary

- No Ultimate Black Box Algorithm™ yet
- Reconstruction quality determined by user choices
- Current challenges:
  - optimal choices of priors and initial image
  - Identification of artefacts
  - simultaneous model fitting
  - optimal reconstruction from multi-frequency datasets
  - reconstruction on non-planar surface (spheroids)