Image Reconstruction in Optical Interferometry

Éric Thiébaut & John Young

Centre de Recherche Astronomique de Lyon
University of Cambridge

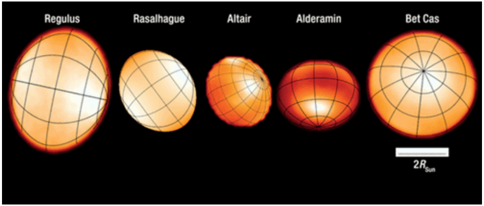
VLTI School – Barcelonnette, 2013

Outline

Motivation for model-independent imaging

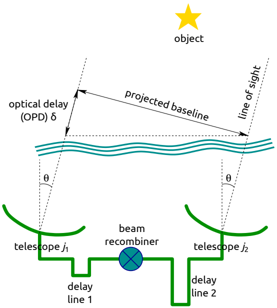
The need for imaging

- Assumptions about basic geometry of an object can be very misleading
  - A model with the wrong geometry can fit well even with moderate uv coverage
  - The best-fit parameters are then completely bogus
- Image reconstruction is often the only reliable way to identify the most appropriate class of models
- Images can be interpreted and analysed straightforwardly by colleagues who are unfamiliar with interferometry
- Images make your results more accessible and improve funding prospects!

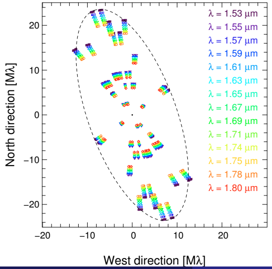


Interferometric Observables

Instantaneous output of an interferometer



$\Longrightarrow$



instantaneous output = *complex visibility*:

$$V_{j_1,j_2}(\lambda,t) = g_{j_1}^{\star}(\lambda,t)\, g_{j_2}(\lambda,t)\, \widehat{I}_{\lambda}(\boldsymbol{b}_{j_1,j_2}(t)/\lambda)$$

with:

- $g_j(\lambda,t)$ = instantaneous complex amplitude transmission for jth telescope;
- $\widehat{I}_{\lambda}(\boldsymbol{\nu})$ = angular Fourier transform of the specific brightness distribution $I_{\lambda}(\boldsymbol{\alpha})$ of the observed object in angular direction $\boldsymbol{\alpha}$;
- projected ***baseline***:

  $$\boldsymbol{b}_{j_1,j_2}(t) = \boldsymbol{r}_{j_2}(t) - \boldsymbol{r}_{j_1}(t)$$

  $\boldsymbol{r}_j(t)$ = position of jth telescope projected on a plane perpendicular to the line of sight;
- λ = wavelength;
- t = time;

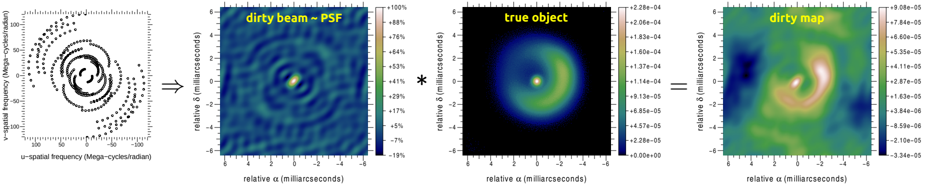
Easy case: image reconstruction $\sim$ deconvolution

At any observed frequency, $\boldsymbol{\nu}_k = \boldsymbol{b}_{j_1,j_2}(t_m)/\lambda_\ell$, the data is given by:

$$z_k = \widehat{h}_k \, \widehat{I}_{\lambda_\ell}(\boldsymbol{\nu}_k) + \text{noise}$$

with the transfer function (the Fourier transform of the *dirty beam*):

$$\widehat{h}_k = g^\star_{j_1}(\lambda_\ell, t_m)\, g_{j_2}(\lambda_\ell, t_m)$$



when the complex visibilities and the complex throughput are available:

image reconstruction $\sim$ deconvolution

⚠ many missing values (very sparse data)

$\Rightarrow$ other constraints (***priors***) than the data are required

The effects of turbulence

Because of the ***atmospheric turbulence***, averaging during an exposure yields:

$$\langle V_{j_1,j_2}(\lambda,t)\rangle_m = \left\langle g^\star_{j_1}(\lambda,t)\, g_{j_2}(\lambda,t)\, \widehat{I}_\lambda(\boldsymbol{\nu}_{j_1,j_2}(\lambda,t))\right\rangle_m$$

$$\approx \underbrace{\langle g_{j_1}(\lambda,t)\rangle^\star_m}_{\approx 0}\, \underbrace{\langle g_{j_2}(\lambda,t)\rangle_m}_{\approx 0}\, \widehat{I}_\lambda(\boldsymbol{b}_{j_1,j_2,m}/\lambda)$$

$\langle\ldots\rangle_m$ means averaging during mth exposure

with: $\boldsymbol{b}_{j_1,j_2,m} \stackrel{\text{def}}{=} \langle \boldsymbol{r}_{j_2}(t)\rangle_m - \langle \boldsymbol{r}_{j_1}(t)\rangle_m$

the mean baseline during the exposure

$\Rightarrow$ we need to integrate observables which are ***insensitive to phase delay errors***:

- ***powerspectrum***

$$\langle |V_{j_1,j_2}(\lambda,t)|^2\rangle_m \approx \underbrace{\langle |g_{j_1}(\lambda,t)|^2\rangle_m\, \langle |g_{j_2}(\lambda,t)|^2\rangle_m}_{>0}\, |\widehat{I}_\lambda(\boldsymbol{b}_{j_1,j_2,m}/\lambda)|^2$$

- ***bispectrum***

$$\langle V_{j_1,j_2}(\lambda,t)\, V_{j_2,j_3}(\lambda,t)\, V_{j_3,j_1}(\lambda,t)\rangle_m \approx \underbrace{\langle |g_{j_1}(\lambda,t)|^2\rangle_m\, \langle |g_{j_2}(\lambda,t)|^2\rangle_m\, \langle |g_{j_3}(\lambda,t)|^2\rangle_m}_{>0}$$

$$\widehat{I}_\lambda(\boldsymbol{b}_{j_1,j_2,m}/\lambda)\, \widehat{I}_\lambda(\boldsymbol{b}_{j_2,j_3,m}/\lambda)\, \widehat{I}_\lambda(\boldsymbol{b}_{j_3,j_1,m}/\lambda)$$

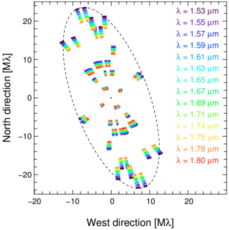
Issues in image reconstruction from optical interferometry data

1. ***sparsity of the data***
   (*holes* in the spatial frequency coverage ▶)
   ⇒ additional ***prior*** needed
2. ***non-linear data***

$$\text{powerspectrum} \propto \left|\widehat{I}_\lambda(\boldsymbol{\nu}_k)\right|^2$$

$$\text{bispectrum} \propto \widehat{I}_\lambda(\boldsymbol{\nu}_{k_1})\,\widehat{I}_\lambda(\boldsymbol{\nu}_{k_2})\,\widehat{I}^\star_\lambda(\boldsymbol{\nu}_{k_1}+\boldsymbol{\nu}_{k_2})$$

3. ***calibration*** of the effective transfer functions
4. ***missing Fourier phases***
   - powerspectrum provides no phase
   - ***phase closure*** (the phase of the bispectrum) only provide 1 phase out of 3



Inverse Approach

Inverse approach for image reconstruction

Inverse approach provides a very general framework to describe most (if not all) image reconstruction algorithms (le Besnerais et al. 2008; Thiébaut 2009; Thiébaut and Giovannelli 2010).

The recipes involve:

1. a **direct model**: model of the brightness distribution and its Fourier transform;
2. a **criterion** to determine a unique and stable solution;
3. an **optimization strategy** to find the solution.

Image and complex visibilities models

Image model

The specific brightness distribution in angular direction $\boldsymbol{\alpha}$ is approximated by:

$$I_\lambda(\boldsymbol{\alpha}) \approx \sum_n b_n(\boldsymbol{\alpha})\, x_n \quad \xrightarrow{\text{F.T.}} \quad \widehat{I}_\lambda(\boldsymbol{\nu}) \approx \sum_n \widehat{b}_n(\boldsymbol{\nu})\, x_n$$

with $\{b_n : \mathbb{R}^2 \mapsto \mathbb{R}\}_{n=1}^N$ a basis of functions and $\boldsymbol{x} \in \mathbb{R}^N$ the ***image parameters***.

Complex visibility model

For any sampled spatial frequency $\boldsymbol{\nu}_k = \boldsymbol{b}_{j_1,j_2,m}/\lambda$ the model complex visibility can be written:

$$\widehat{I}_\lambda(\boldsymbol{\nu}_k) \approx y_k = \sum_n \widehat{b}_n(\boldsymbol{\nu}_k)\, x_n = \sum_n H_{k,n}\, x_n$$

with $H_{k,n} = \widehat{b}_n(\boldsymbol{\nu}_k)$, in matrix notation:

$$\boldsymbol{y} = \mathbf{H} \cdot \boldsymbol{x}$$

with $\boldsymbol{y} \in \mathbb{C}^K$ and $\mathbf{H} \in \mathbb{C}^{K \times N}$ is a sub-sampled Fourier transform operator.

Image constraints

Image reconstruction is a compromise between various constraints (Thiébaut 2009).

Data constraints

The image must be ***compatible with the data*** z (powerspectrum, bispectrum, *etc.*):

$$f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}) \stackrel{\mathsf{def}}{=} -\log \mathrm{pdf}(\boldsymbol{z}|\mathbf{H}\cdot\boldsymbol{x}) + c \leq \eta$$

with $\mathrm{pdf}(\boldsymbol{z}|\mathbf{H}\cdot\boldsymbol{x})$ the ***likelihood*** of the data given the model and $\eta > 0$.

⚠ Even with $\eta = 0$, this is insufficient to define a unique (and stable) solution, we need additional ***a priori constraints***:

Strict priors

e.g. the image must be non-negative and normalized

$$\forall n, x_n \geq 0 \quad \text{and} \quad \sum_n x_n = 1$$

$$\iff \quad \boldsymbol{x} \in \mathbb{X} \qquad \text{(the feasible set)}$$

Loose priors

e.g. the image must be *simple* or *smooth*

$$\min_{\boldsymbol{x}\in\mathbb{X}} f_{\mathsf{prior}}(\boldsymbol{x})$$

Inverse problem formulation

We want to follow the priors as far as possible providing the image remains compatible with the data:

$$\boldsymbol{x}^{+} = \underset{\boldsymbol{x}\in\mathbb{X}}{\arg\min}\, f_{\text{prior}}(\boldsymbol{x}) \quad \text{s.t.} \quad f_{\text{data}}(\mathbf{H}\cdot\boldsymbol{x}) \leq \eta$$

which can be solved via the Lagrangian:

$$\mathcal{L}(\boldsymbol{x};\ell) = f_{\text{prior}}(\boldsymbol{x}) + \ell\, f_{\text{data}}(\mathbf{H}\cdot\boldsymbol{x})$$

with $\ell \geq 0$ the Lagrange multiplier for the inequality constraint $f_{\text{data}}(\mathbf{H}\cdot\boldsymbol{x}) \leq \eta$. The inequality constraint must be active, hence $\ell > 0$ and, taking $\mu = 1/\ell$, leads to the solution:

Maximum a posteriori solution

$$\boldsymbol{x}(\mu) = \underset{\boldsymbol{x}\in\mathbb{X}}{\arg\min}\, f(\boldsymbol{x};\mu)$$

$$\text{with:} \quad f(\boldsymbol{x};\mu) = \underbrace{f_{\text{data}}(\mathbf{H}\cdot\boldsymbol{x})}_{\text{likelihood}} + \underbrace{\mu\, f_{\text{prior}}(\boldsymbol{x})}_{\text{regularization}}$$

where $\mu > 0$ is tuned so as to match $f_{\text{data}}(\mathbf{H}\cdot\boldsymbol{x}^{+}) = \eta$ with $\boldsymbol{x}^{+} = \boldsymbol{x}(\mu^{+})$ and μ^{+} the optimal regularization weight.

Likelihood of the Data

Likelihood of the data

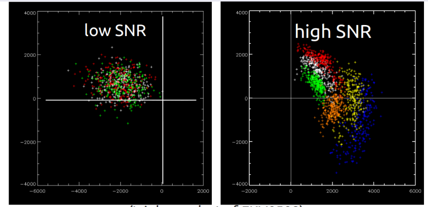
- should be based on the noise statistics of the data:

$$f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}) \stackrel{\text{def}}{=} -\log \mathrm{pdf}(\boldsymbol{z}|\mathbf{H}\cdot\boldsymbol{x}) + c$$

- can be very complicated (non-convex, phase wrapping, *etc.*)
- various approximations have been proposed (*e.g.*, Meimon et al. 2005a)

⚠ in general this does not amounts to least-squares (even weighted ones!)

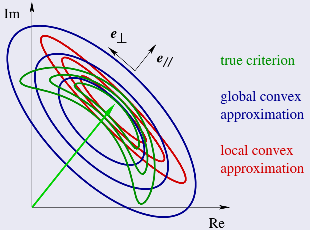
Real complex data



(triple product of FKV0509)

Source : C. Hummel *et al.* <http://www.mrao.cam.ac.uk/~jsy1001/exchange/complex/complex.html>

Approximate cost function

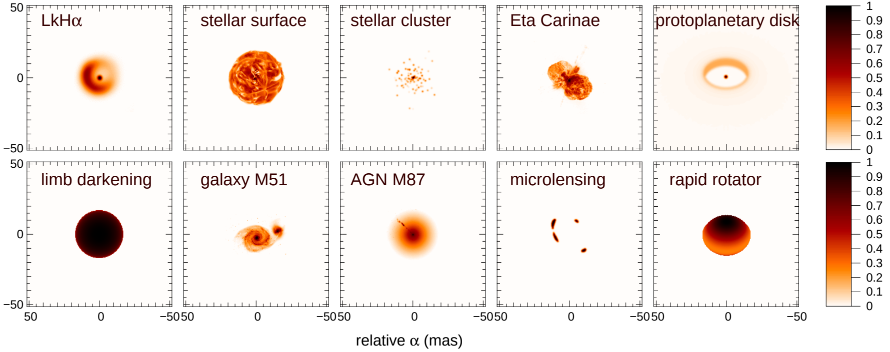


Regularization

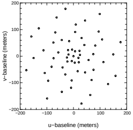
Which are the best regularization methods?

Practical comparison of regularization methods:

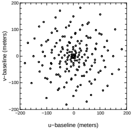
- a study made by S. Renard *et al.* (Astron. & Astrophys., 2011);
- about **20 000** simulations:
  - 10 different objects;
  - 11 different regularizations;
  - 20 regularization levels;
  - 3 different (u, v) coverages: *poor* (31 freq.), *medium* (88 freq.), and *rich* (245 freq.);
  - 3 different signal-to-noise ratii (SNR): *high* (1%), *medium* (5%), and *low* (10%);
- assumptions: complex visibilities available
  $\Longrightarrow$ ***convex*** constrained non-linear optimization problem;
- algorithm: ***MiRA*** (Thiébaut, 2008, 2009);



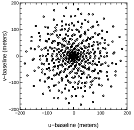
Various objects ▲ and (u, v) coverages ▶ (Renard et al. 2011).



31 freq.



88 freq.



245 freq.

Various regularizations

We consider the following regularizations:

1. **Quadratic smoothness**:

$$f_{\text{prior}}(\boldsymbol{x}) = \|\boldsymbol{x} - \mathbf{S}\cdot\boldsymbol{x}\|^2$$

where $\mathbf{S}$ is a smoothing operator (by finite differences).

2-3. **Compactness** (le Besnerais et al. 2008):

$$f_{\text{prior}}(\boldsymbol{x}) = \sum_n w_n^{\text{prior}} \boldsymbol{x}_n^2$$

with $w_n^{\text{prior}} = \|\boldsymbol{\theta}_n\|^\beta$ and $\beta = 2$ or 3 yields ***spectral smoothness***.

4-5. **Non-linear smoothness:**

$$f_{\text{prior}}(\boldsymbol{x}) = \sum_n \sqrt{\|\nabla x_n\|^2 + \epsilon^2}$$

where $\|\nabla x_n\|^2$ is the squared magnitude of the spatial gradient in the image at nth pixel and $\epsilon \to 0$ yields **total variation** (Rudin et al. 1992) while $\epsilon > 0$ yields **edge-preserving smoothness** (Charbonnier et al. 1997).

Various regularizations (continued)

6-8. **Separable norms (ℓ_p):**

$$f_{\text{prior}}(\boldsymbol{x}) = \sum_n \left(x_n^2 + \epsilon^2\right)^{p/2} \approx \sum_n |x_n|^p$$

where $\epsilon > 0$ and $p = 1.5$, 2, and 3. Note that $p = 1$ is what is advocated in ***compress sensing*** (Donoho 2006) while $p = 2$ corresponds to regular ***Tikhonov regularization***.

9-11. **Maximum entropy methods** (Narayan and Nityananda 1986):

$$f_{\text{prior}}(\boldsymbol{x}) = -\sum_n h(x_n; \bar{x}_n).$$

Here the prior is to assume that the image is drawn toward a prior model $\bar{\boldsymbol{x}}$ according to a non quadratic potential h, called the ***entropy***:

$$\begin{aligned}
&\text{MEM-sqrt:} && h(x;\bar{x}) = \sqrt{x}\,; \\
&\text{MEM-log:} && h(x;\bar{x}) = \log(x)\,; \\
&\text{MEM-prior:} && h(x;\bar{x}) = x - \bar{x} - x\,\log\left(x/\bar{x}\right)\,.
\end{aligned}$$

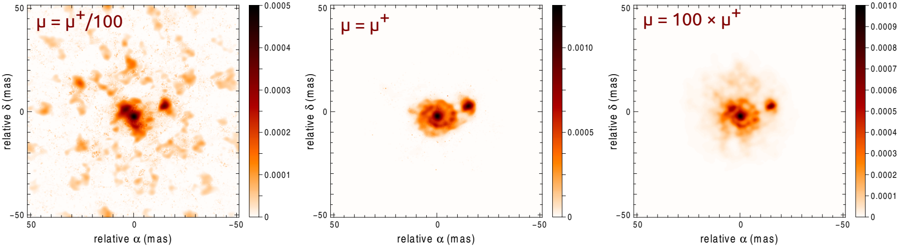
Tuning the regularization level

We choose the regularization level μ^+ by minimizing the ***mean squared error*** (MSE) of the reconstruction versus the true image:

$$\mu^+ = \arg\min_{\mu > 0} \left\| \boldsymbol{x}(\mu) - \boldsymbol{x}^{\mathsf{true}} \right\|_2$$

where

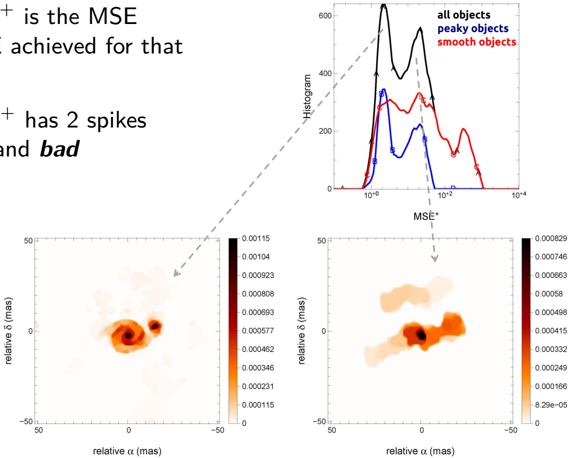
$$\boldsymbol{x}(\mu) \stackrel{\mathsf{def}}{=} \arg\min_{\boldsymbol{x} \in \mathbb{X}} \left\{ f_{\mathsf{data}}(\mathbf{H} \cdot \boldsymbol{x}) + \mu \, f_{\mathsf{prior}}(\boldsymbol{x}) \right\}$$



Is the MSE$^+$ a good figure of merit?

For a given object, MSE$^+$ is the MSE divided by the best MSE achieved for that object.

The distribution of MSE$^+$ has 2 spikes corresponding to ***good*** and ***bad*** reconstructions.



And the winner is...

Cumulative Performance Rank

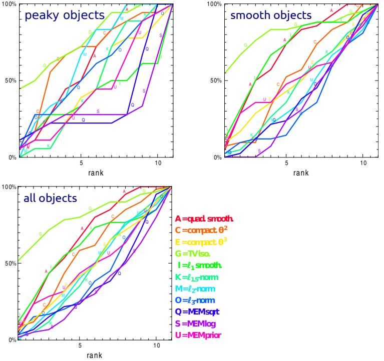
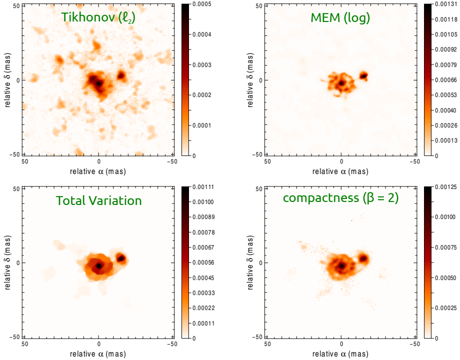


Image Reconstructions



Based on cumulative rank, ***TV*** and ***compactness*** are the most successful.

However the best prior depends on the particular case (object type, SNR and coverage).

Optimization Strategy

Image reconstruction = optimization problem

Assuming $\mu^+ = 1$, image reconstruction amounts to solve:

$$\min_{\boldsymbol{x}\in\mathbb{X}} \underbrace{\{f_{\mathsf{prior}}(\boldsymbol{x}) + f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x})\}}_{f(\boldsymbol{x})}$$

For optical interferometric data, the joint criterion $f(\boldsymbol{x})$ is:

- highly non-linear (means *non-quadratic*);
- depending on a very large number of parameters (the image pixels);
- multimodal $\Longrightarrow$ in principle, needs *global optimization* or a good starting point followed by continuous optimization;

Proposed methods:

- matching-pursuit: ***CLEAN*** (Fomalont 1973; Högbom 1974), the ***building-blocks*** method (Hofmann and Weigelt 1993)
- self-calibration: ***Wisard*** (Meimon et al. 2005b);
- direct optimization: ***BSMEM*** (Baron and Young 2008), ***MiRA*** (Thiébaut 2008);
- global optimization: ***MACIM*** (Markov Chain Imager, Ireland et al. 2008);

Self-calibration

Self-calibration (Readhead and Wilkinson 1978; Schwab 1980; Cornwell and Wilkinson 1981) proposed to solve for missing calibration of the transfer function or missing Fourier phases.

Self-calibration algorithm

Choose an initial image $\boldsymbol{x}^{[0]}$ and repeat the following steps for $k = 0, 1, \ldots$ until convergence:

1. **self-calibration** step:

$$\boldsymbol{y}^{[k+1]} = \arg\min_{\boldsymbol{y}} f_{\text{data}}(\boldsymbol{y}) \quad \text{s.t.} \quad \boldsymbol{y} \approx \mathbf{H} \cdot \boldsymbol{x}^{[k]}$$

2. **image reconstruction** step (deconvolution):

$$\boldsymbol{x}^{[k+1]} = \arg\min_{\boldsymbol{x} \in \mathbb{X}} f_{\text{prior}}(\boldsymbol{x}) \quad \text{s.t.} \quad \mathbf{H} \cdot \boldsymbol{x} \approx \boldsymbol{y}^{[k+1]}$$

Issues:

- What is the meaning of $\approx$ (depends on the algorithm)?
- How to consistently tune the balance between prior and data?
- Not rigorously equivalent to minimizing a given criterion.

Augmented Lagrangian approach

Solving the image reconstruction problem by ***direct minimization*** of the criterion, *i.e.*

$$\min_{\boldsymbol{x}\in\mathbb{X}} \{ f_{\mathsf{prior}}(\boldsymbol{x}) + f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}) \}$$

is exactly the same as solving the ***constrained problem***:

$$\min_{\boldsymbol{x}\in\mathbb{X},\boldsymbol{y}} \{ f_{\mathsf{prior}}(\boldsymbol{x}) + f_{\mathsf{data}}(\boldsymbol{y}) \} \quad \text{s.t.} \quad \mathbf{H}\cdot\boldsymbol{x} = \boldsymbol{y}$$

where the ***model complex visibilities*** $\boldsymbol{y} = \mathbf{H}\cdot\boldsymbol{x}$ have been explicitly introduced as ***auxiliary variables***.

The ***augmented Lagrangian*** (Boyd et al. 2010) is a practical algorithm to solve this constrained problem:

$$\mathcal{L}_{\mathsf{A}}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u};\beta) = f_{\mathsf{prior}}(\boldsymbol{x}) + f_{\mathsf{data}}(\boldsymbol{y}) - \boldsymbol{u}^{\mathrm{T}}\cdot[\mathbf{H}\cdot\boldsymbol{x} - \boldsymbol{y}] + \frac{\beta}{2}\,\|\mathbf{H}\cdot\boldsymbol{x} - \boldsymbol{y}\|^2 ,$$

with $\boldsymbol{u}$ the Lagrange multipliers related to the constraints $\mathbf{H}\cdot\boldsymbol{x} = \boldsymbol{y}$ and $\beta > 0$ the weight of the quadratic penalty to reinforce the constraints.

Advantages: explicit update formula for the Lagrange multipliers, strong convergence properties for β large enough (no need for $\beta \to \infty$), *etc.*

Augmented Lagrangian approach (continued)

$$\mathcal{L}_{\mathsf{A}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}; \beta) = f_{\mathsf{prior}}(\boldsymbol{x}) + f_{\mathsf{data}}(\boldsymbol{y}) - \boldsymbol{u}^{\mathrm{T}} \cdot [\mathbf{H} \cdot \boldsymbol{x} - \boldsymbol{y}] + \frac{\beta}{2} \|\mathbf{H} \cdot \boldsymbol{x} - \boldsymbol{y}\|^2$$

Augmented Lagrangian algorithm (in our case)

Start with initial multipliers $\boldsymbol{u}^{[0]}$ and $\beta^{[0]} > 0$ and repeat the following steps for $k = 0, 1, \ldots$ until convergence:

1. improve the variables:

$$\{\boldsymbol{x}, \boldsymbol{y}\}^{[k+1]} \approx \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{X}, \boldsymbol{y}} \mathcal{L}_{\mathsf{A}} \left(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}^{[k]}; \beta^{[k]}\right)$$

2. update the multipliers:

$$\boldsymbol{u}^{[k+1]} = \boldsymbol{u}^{[k]} + \beta \left(\boldsymbol{y}^{[k+1]} - \mathbf{H} \cdot \boldsymbol{x}^{[k+1]}\right)$$
$$\beta^{[k+1]} = \beta^{[k]}$$

or strengthen the constraints:

$$\boldsymbol{u}^{[k+1]} = \boldsymbol{u}^{[k]}$$
$$\beta^{[k+1]} = \gamma\, \beta^{[k]} \qquad \text{(with } \gamma > 1\text{)}$$

Step 1 can be implemented thanks to alternating minimization, *e.g.*:

$$\boldsymbol{x}^{+} = \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{X}} \mathcal{L}_{\mathsf{A}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}; \beta) \quad \text{followed by} \quad \boldsymbol{y}^{+} = \operatorname*{arg\,min}_{\boldsymbol{y}} \mathcal{L}_{\mathsf{A}}(\boldsymbol{x}^{+}, \boldsymbol{y}, \boldsymbol{u}; \beta)$$

Image reconstruction step in augmented Lagrangian approach

The augmented Lagrangian can be rewritten as:

$$\begin{aligned}\mathcal{L}_{\mathsf{A}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}; \beta) &= f_{\mathsf{prior}}(\boldsymbol{x}) + f_{\mathsf{data}}(\boldsymbol{y}) - \boldsymbol{u}^{\mathrm{T}} \cdot [\mathbf{H} \cdot \boldsymbol{x} - \boldsymbol{y}] + \frac{\beta}{2} \|\mathbf{H} \cdot \boldsymbol{x} - \boldsymbol{y}\|^2 , \\ &= f_{\mathsf{prior}}(\boldsymbol{x}) + f_{\mathsf{data}}(\boldsymbol{y}) + \frac{\beta}{2} \|\mathbf{H} \cdot \boldsymbol{x} - \boldsymbol{y} - \boldsymbol{u}/\beta\|^2 - \frac{1}{2\,\beta} \|\boldsymbol{u}\|^2 .\end{aligned}$$

Improving $\boldsymbol{x}$ given the other variables writes:

$$\begin{aligned}\boldsymbol{x}^{+} &= \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{X}} \mathcal{L}_{\mathsf{A}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}; \beta) \\ &= \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{X}} \left\{ f_{\mathsf{prior}}(\boldsymbol{x}) + \frac{\beta}{2} \|\mathbf{H} \cdot \boldsymbol{x} - \boldsymbol{v}\|^2 \right\} \quad \text{with} \quad \boldsymbol{v} = \boldsymbol{y} + \boldsymbol{u}/\beta .\end{aligned}$$

which is the analogous of ***image reconstruction*** given *pseudo-complex visibilities* $\boldsymbol{v} = \boldsymbol{y} + \boldsymbol{u}/\beta$ with white noise of variance $\propto \beta^{-1/2}$ (unlike self-calibration which would try to fit $\boldsymbol{y}$).

Calibration step in augmented Lagrangian approach

Recalling that the augmented Lagrangian can be rewritten as:

$$\mathcal{L}_{\mathsf{A}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}; \beta) = f_{\mathsf{prior}}(\boldsymbol{x}) + f_{\mathsf{data}}(\boldsymbol{y}) + \frac{\beta}{2} \left\| \mathbf{H} \cdot \boldsymbol{x} - \boldsymbol{y} - \boldsymbol{u}/\beta \right\|^2 - \frac{1}{2\,\beta} \left\| \boldsymbol{u} \right\|^2 ,$$

improving $\boldsymbol{y}$ given the other variables writes:

$$\begin{aligned} \boldsymbol{y}^{+} &= \arg\min_{\boldsymbol{y}} \mathcal{L}_{\mathsf{A}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}; \beta) \\ &= \arg\min_{\boldsymbol{y}} \left\{ f_{\mathsf{data}}(\boldsymbol{y}) + \frac{\beta}{2} \left\| \boldsymbol{y} - \boldsymbol{w} \right\|^2 \right\} \quad \text{with} \quad \boldsymbol{w} = \mathbf{H} \cdot \boldsymbol{x} - \boldsymbol{u}/\beta \,. \end{aligned}$$

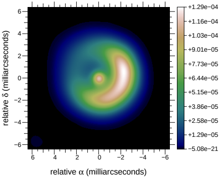
which is similar to the self-calibration step in self-calibration methods except that the complex visibilities $\boldsymbol{y}$ are enforced to fit the actual data and the *shifted* model complex visibilities $\boldsymbol{w} = \mathbf{H} \cdot \boldsymbol{x} - \boldsymbol{u}/\beta$ and not just $\mathbf{H} \cdot \boldsymbol{x}$.

Conclusions about optimization strategy

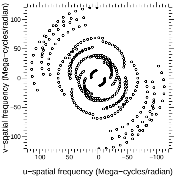
- ***direct optimization*** is more consistent (the given criterion is minimized) and much faster and stable than *self-calibration* for finding missing Fourier phases (as in Wisard, Meimon et al. 2005b) or missing parameters in the OTF:
  - imposing $\boldsymbol{u} = \mathbf{0}$ for the Lagrange multipliers yields the same method as *self-calibration*;
  - exactly matching $\mathbf{H} \cdot \boldsymbol{x} = \boldsymbol{y}$ requires $\beta \to \infty$ which worsens the condition number of the problem and, thus slows down convergence;
  - direct optimization is more consistent (the given criterion is minimized) and much faster and stable;
- ***direct optimization*** with ℓ_1 regularization (to impose sparsity) is superior to *matching pursuit* (Marsh and Richardson 1987) for imposing the sparsity in the CLEAN (Fomalont 1973; Högbom 1974) and *building-blocks* (Hofmann and Weigelt 1993) methods;
- the most successful algorithms – *e.g.* BSMEM (Baron and Young 2008) and MiRA (Thiébaut 2008) – use direct optimization;

global optimization is however required, *e.g.* attempt by the Markov Chain Imager (MACIM) algorithm (Ireland et al. 2008);

Example of Image Reconstruction



true object (smoothed)



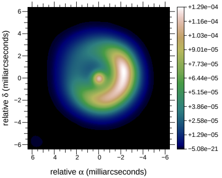
u-v coverage

- simulated data for *Beauty Contest 2004* (Lawson *et al.* 2004)
- reconstruction by MiRA algorithm (Thiébaut, 2008)

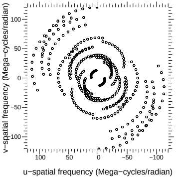
$$\boldsymbol{x} = \underset{\boldsymbol{x}\in\mathbb{X}}{\arg\min}\, f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}) + \mu\, f_{\mathsf{prior}}(\boldsymbol{x})$$

constrained non-linear optimization by limited memory quasi-Newton method

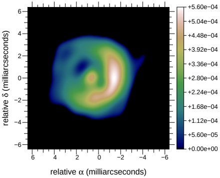
Example of Image Reconstruction



true object (smoothed)



u-v coverage



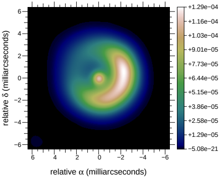
reconstruction with powerspectrum and phase closures

- simulated data for *Beauty Contest 2004* (Lawson *et al.* 2004)
- reconstruction by MiRA algorithm (Thiébaut, 2008)

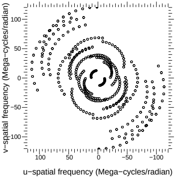
$$\boldsymbol{x} = \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{X}} f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}) + \mu\, f_{\mathsf{prior}}(\boldsymbol{x})$$

constrained non-linear optimization by limited memory quasi-Newton method

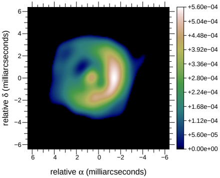
Example of Image Reconstruction



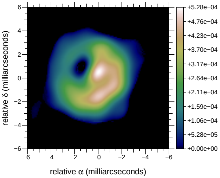
true object (smoothed)



u-v coverage



reconstruction with powerspectrum and phase closures



reconstruction with powerspectrum only

- simulated data for *Beauty Contest 2004* (Lawson *et al.* 2004)
- reconstruction by MiRA algorithm (Thiébaut, 2008)

$$\boldsymbol{x} = \underset{\boldsymbol{x} \in \mathbb{X}}{\arg\min}\, f_{\mathsf{data}}(\mathbf{H} \cdot \boldsymbol{x}) + \mu\, f_{\mathsf{prior}}(\boldsymbol{x})$$

constrained non-linear optimization by limited memory quasi-Newton method

Existing Algorithms

Algorithm Ingredients

- Supported data types and corresponding likelihood functions
- Image model and method for forward transform to data space (= ***direct model***)
- Strict constraints (positivity and normalization of image)
- Prior (***regularization***) type and level
  - Possible prior model
- Algorithm for solving the inverse problem
  - How the inverse problem is expressed
  - Numerical algorithm used to solve it
  - Starting model

Algorithm Comparison

Name	Authors	Optimization	Regularization
BSMEM	Baron, Buscher	Trust region gradient	MEM-prior
MiRA	Thiébaut	VMLM-B$^{(\star)}$	Many
WISARD	Meimon, Mugnier, Le Besnerais	VMLM-B$^{(\star)}$ plus self-calibration	Many
MACIM	Ireland, Monnier	Simulated annealing	MEM
SQUEEZE	Baron, Monnier, Kloppenborg	Parallel tempering	
Building Block method	Hofmann, Weigelt	Matching pursuit	Sparsity

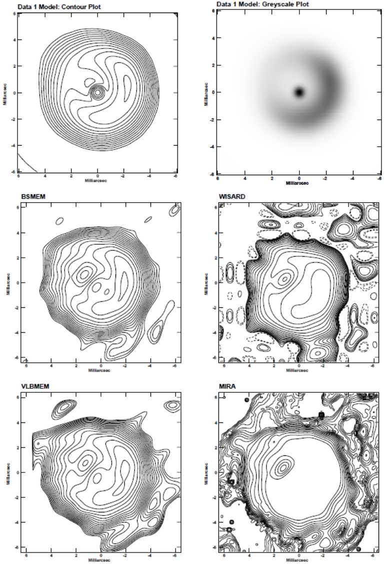
$^{(\star)}$ VMLM-B is a quasi-Newton method with bounds on the parameters (Thiébaut 2002)

Points of difference

- Image model
  - Conventional grid of pixels
  - Sparsity basis (compressed sensing) – work in progress
  - Fourier transform implementation (handling of uneven Fourier sampling)
- Treatment of the observables
  - Explicit solving for phases (WISARD)
  - Direct use of OI observables (e.g. MiRA, BSMEM), locally convex likelihood
  - Noise model for complex quantities (c.f. OIFITS standard)
- Bayesian/non-Bayesian algorithm
  - Stopping criterion
  - Treatment of hyperparameter
  - Evidence evaluated?
- Global or gradient optimization
  - Gradient optimization needs differentiable regularizer
  - Global optimization by Markov Chain Monte-Carlo techniques
- Available regularizers
- User interface
- Availability of the code, documentation and support

Results

- Despite algorithm differences, usually get very similar results!
    - Note importance of strict constraints (MiRA image was reconstructed *without* normalization)
- For this dataset
    - All algorithms recover the correct morphology
    - All get the astrometry and photometry wrong. . .



Are there sufficient data for image reconstruction?

Enough data?

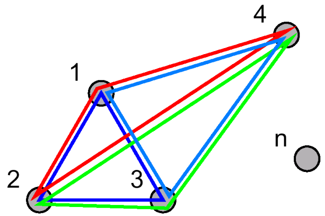
- The number of independent uv points $\geq$ number of filled resolution elements in the recovered image
  - Don't bother trying with < 20 data
- The range of interferometer baselines i.e. $B_{\max}/B_{\min}$ will govern the range of spatial scales in the image
  - Need two-dimensional uv coverage
  - Shortest baseline should be well inside the first lobe of the visibility function
- Holes in the uv coverage will give artefacts in the image

Fraction of phase information

$$\text{No. of C.P.} = \frac{(N)(N-1)(N-2)}{(3)(2)}$$

$$\text{No. of indep. C.P.} = \frac{(N-1)(N-2)}{2}$$

$$\text{No. of Phases} = \frac{(N)(N-1)}{2}$$



Φ (1-2-3) = Φ (1-2-4) + Φ (4-2-3) + Φ (1-4-3)

In General:

Φ (1-2-3) = Φ (1-2-n) + Φ (n-2-3) + Φ (1-n-3)

- 3 telescopes $\Longrightarrow$ 1/3 phase information;
  8 telescopes $\Longrightarrow$ 75% phase information
- Impact on reconstruction depends on object morphology – e.g. missing phases have little impact for symmetric objects

Image Reconstruction Parameters

Recap of Parameters

- OIFITS file
- Data selection parameters: observing target, wavelength and time ranges
- Image model
- Regularization type and parameters
  - May include a prior model (required by MEM-prior)
  - Hyperparameter μ (relative weighting of likelihood and prior, may be determined automatically)
- Optimization parameters
  - Starting model
  - Positivity and normalization constraints usually applied by default

Choosing Image Model Parameters

- Usually specify the image dimensions (e.g. 128×128) and pixel scale (e.g. 0.1 mas/pixel)
- These should reflect the range of spatial frequencies in the data and the maximum size of the object
- Pixel size 0.1–0.2 $\lambda_{\min}/B_{\max}$ – algorithms can give super-resolution beyond λ/B
  - Degree of super-resolution depends on noise level and uv coverage of data
- Image width $\gtrsim$ 2–3 $\times$ object size, to avoid aliasing

Choosing the Regularization

- Total variation and compactness are the most successful over a wide range of objects (Renard et al. 2011)
- MiRA supports a wide range of priors, including user-defined ones, whereas BSMEM only supports MEM-prior
- Regularization terms can, in principle, be calculated with respect to a prior model (default model) – required for MEM-prior
  - Non-flat prior model fixes the position of the object, which is unconstrained if only amplitude/closure phase data
  - Otherwise starting model can be used to enforce the object position (MiRA)
  - Informative prior models can be especially useful for sparse and/or noisy data; otherwise they just speed up convergence
- Hyperparameter μ controls the relative weighting of the likelihood and prior
  - Can be determined objectively by evaluating the Bayesian evidence – BSMEM does this
  - Table of empirical values in Renard et al. (2011)
  - Can try a range of values and select the one that gives $f_{\text{data}} \sim m$

Example Image Reconstruction Sessions

BSMEM

```
jsy1001@cstdev: ~
jsy1001@cstdev:~$
jsy1001@cstdev:~$ bsmem -h


**********  BSMEM v1.5   ******************
Usage: bsmem -d OIFITSfile [-f outputimagefile -p pixellation -w imagewidth ...]

Example: './bsmem -d data.oifits -p 0.1 -w 128'

-h:                Display this information.
-d:                OIFITS file containing the visibility data.
-f:                FITS file to output the reconstructed image.
-sf:               Starting image or prior file. Overrides the -mt command.
-mt:               Model/prior image type.
                     0 : Flat prior.
                     1 : Dirac, centered in the FOV.
                     2 : Uniform disk.
                     3 : Gaussian.
                     4 : Lorentzian.
-mw:               Model witdth (Gaussian and Uniform Disk only).
-mf:               Total flux of the model.
-p:                Size of a pixel (in mas). Set to 0 for automatic.
-w:                Width (in pixels) of the reconstructed image.
-e:                Entropy functional.
```

Useful BSMEM Options

- Specify data file: `-d data.oifits`
- Override default image size and automatic pixel scale: `-w 128 -p 0.2`
- Specify model (used as prior model and starting model):
  - e.g. 20 mas radius Uniform disk: `-mt 2 -mw 20.0`
  - e.g. 30 mas FWHM Gaussian: `-mt 3 -mw 30.0`
- Alternative – specify model image file: `-sf model.fits`
  - Image dimensions and pixel scale must match BSMEM options
- Perhaps adjust error on zero-baseline powerspectrum point: `-ferr 1e-3`
- If extra iterations needed: `-it 400` or `-it -1` (unlimited)
- If scripting BSMEM disable command prompt, specify wavelengths and (optionally) output file: `-noui -wavmin 1680.0 -wavmax 1720.0 -f out.fits`
- Alternative – redirect commands from file to stdin:

```
bsmem -d data.fits < bsmem.in
```

Starting BSMEM (i)

```
jsy1001@cstdev: ~/Dropbox/STORE/Simulations/OI_Imaging/VLTI_School_2013
**********  BSMEM v1.5   ******************
Datafile:               RSG_distX7_H_MRO8bs_D_sy123.oifits
Reading unit labels:    OI_TARGET OI_WAVELENGTH OI_VIS2 OI_VIS2 OI_VIS2 OI_T3 OI
_T3 OI_T3
Target id/name:         1/Fake_Targ
Auto selecting the only target "Fake_Targ".

POWERSPECTRUM TABLES
#       Date            Array                   Instrument              Nrec/Nwa
v
001     2009-08-06                              Fake_Ins                285/5
002     2009-08-06                              Fake_Ins                285/5
003     2009-08-06                              Fake_Ins                285/5

BISPECTRUM TABLES
#       Date            Array                   Instrument              Nrec/Nwa
v
001     2009-04-15                              Fake_Ins                190/5
002     2009-04-15                              Fake_Ins                190/5
003     2009-04-15                              Fake_Ins                190/5

INSTRUMENT SPECTRAL CHANNELS
#       Instrument              Channel_id      Band/Bandwidth (nm)
0       Fake_Ins                 000_000                1540/55
```

Starting BSMEM (ii)

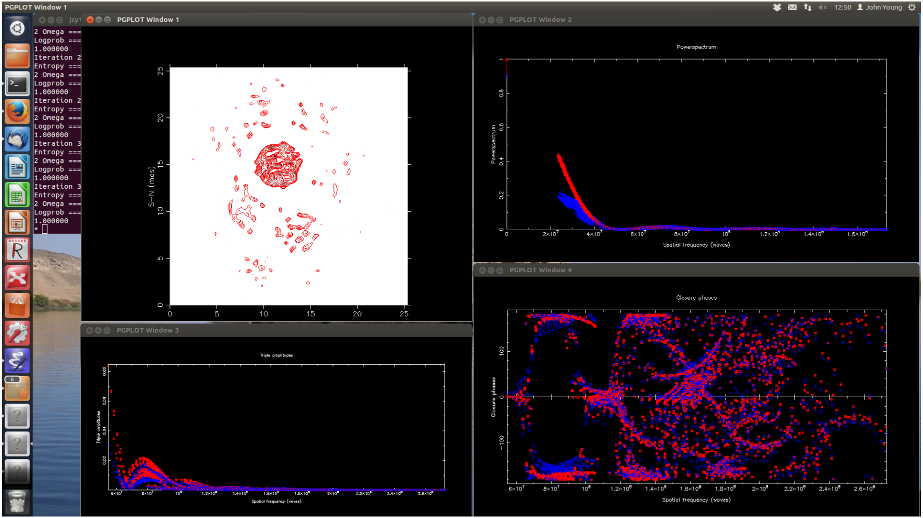
```
jsy1001@cstdev: ~/Dropbox/STORE/Simulations/OI_Imaging/VLTI_School_2013
                                005_002                 1650/55
                                005_003                 1705/55
                                005_004                 1760/55
Select a wavelength range (default value = 1 50000) :1640 1660
Found 855 powerspectrum and 570 bispectrum points between 1640 and 1660 nm.

Bispectrum noise:        Classic elliptic approximation
UV range:                25022904 - 161907664 wavelengths
Array resolution:         0.636983 mas
Pixel size:              Automatic, 0.212328 mas
Recommended size:         64 pixels
Image width:             128 pixels, 27.177935 mas
Pix/fastest fringe:      6.000000
Entropy functional:      Gull-Skilling entropy
Hyperparameter scheme:   Chi2 = N method
Maximum n# iterations:   200
Gaussian, FWHM:10.000000 mas, sigma:4.246609 mas, flux:0.010000

Starting Maximum Entropy Reconstruction.
Iteration 1 Ntrans === 4 istat === 0101000
Entropy === 0.000000  Chisq === 1926.560181 Flux === 0.010000 Alpha === 18.23840
9 Omega === 0.000519
Logprob === 0.000000 Good Measurements === 0.000000 Scale === 1.000000
*
```

Useful BSMEM Commands

Commands are not case sensitive.

`EXIT`	
`DO n`	Iterate (`DO -1` to convergence)
`SCALE x`	Set image display exponent
`REDISP ON` / `REDISP OFF`	Enable/disable image and graphs
`CENTER`	Re-centre image
`SNR`	Display signal-to-noise
`UV`	Plot uv-plane coverage
`WRITEFITS`	Save reconstructed image

BSMEM Graphs



Simple MiRA session

0. Start MiRA: launch Yorick and
```
include, "mira.i";
```
1. Load input data into opaque object db:
```
db = mira_new("data/beauty-2004-data1.oifits");
```
2. Configure for image reconstruction:
```
mira_config, db, xform="nfft", dim=150,
    pixelsize=0.1*MIRA_MILLIARCSECOND;
```
3. Choose a regularization method:
```
rgl = rgl_new("smoothness");
```
4. Attempt an image reconstruction (from scratch):
```
dim = mira_get_dim(db);
img0 = array(double, dim, dim);
img0(dim/2, dim/2) = 1.0;
img1 = mira_solve(db, img0, maxeval=500, verb=1, xmin=0.0,
    normalization=1, regul=rgl, mu=1e6);
```
5. Continue reconstruction with recentered image:
```
img1 = mira_solve(db, mira_recenter(img1), maxeval=500,
    verb=1, xmin=0.0, normalization=1, regul=rgl, mu=1e6);
```

Useful MiRA options

Useful options of `mira_solve`:

- `xmin=0.0` to enforce positivity
- `normalization=1` to enforce normalization
- `regul=..., mu=...` to specify regularization type and level
- `verb=`n verbose every n iteration / quiet with `verb=0`
- `maxeval=...` to set maximum number of evaluations

MiRA session with another regularization

0. Start Mira
1. Load input data into opaque object db
2. Configure for image reconstruction
3. Choose a regularization method:

```
dim = mira_get_dim(db);
img0 = array(double, dim, dim);
img0(dim/2, dim/2) = 1.0;
rgl = rgl_new("totvar", epsilon=1e-4, isotropic=1);
```

4. Attempt an image reconstruction (from scratch):

```
img1 = mira_solve(db, img0, maxeval=500, verb=1,
    xmin=0.0, normalization=1, regul=rgl, mu=1e6);
```

5. Change a regularisation parameter and continue reconstruction with recentered image:

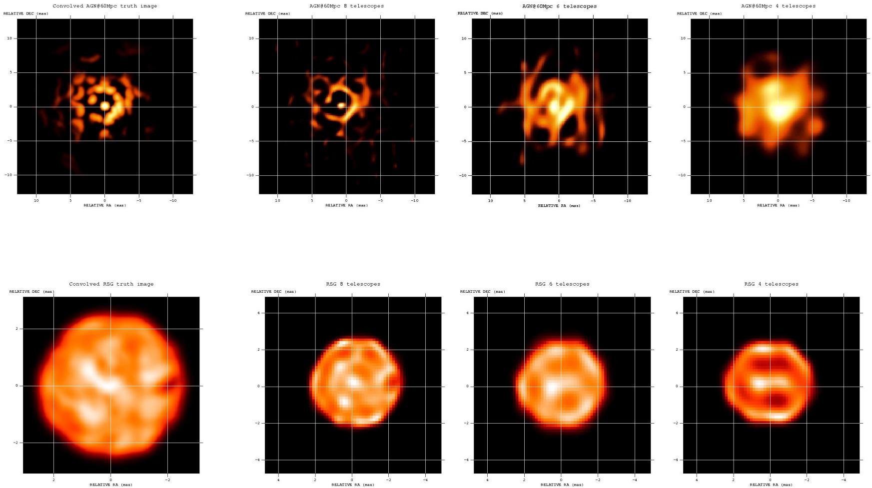
```
rgl_config, rgl, epsilon=1e-3;
img1 = mira_solve(db, mira_recenter(img1), maxeval=500,
    verb=1, xmin=0.0, normalization=1,
    regul=rgl, mu=1e4);
```

Interpretation of reconstructed images

Can we improve the reconstruction?

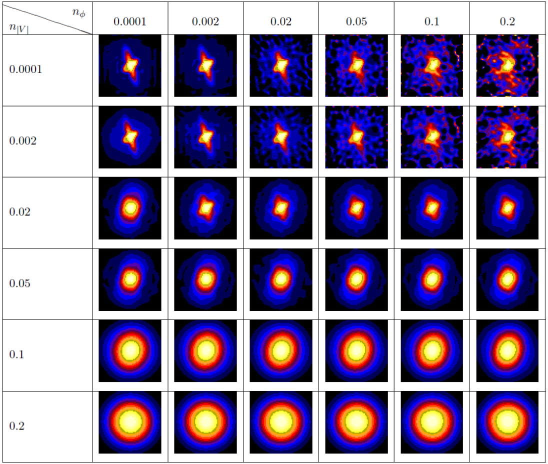
- Adjust the size of the support to the reconstructed object
  - Map size
  - Width of prior model
- Use the reconstructed object to inform the choice of prior/starting model
  - Beneficial at intermediate SNR or if uv coverage poor
  - Can use initial reconstruction, thresholded and smoothed, as model for second run
- Re-center the object part-way through the reconstruction
- Experiment with selected wavelength range
  - Trade improved uv coverage against intrinsic variation of object with wavelength
- Experiment with selected timespan
  - Trade improved uv coverage against intrinsic variation of object with time

Effect of uv coverage



All simulations are of 6 hour observations

Effect of signal-to-noise



Simulations from (Baron, 2007, BSMEM report)

What features are believable?

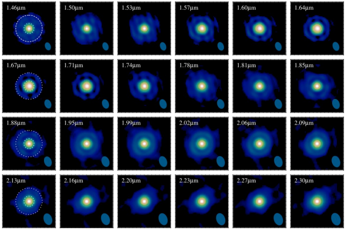
Usually difficult to identify a “noise level” in the reconstructed image, due to regularization and artefects of sparse uv coverage.
Instead we must consider:

- Are features robust to changing reconstruction parameters?
- Compare reconstructions from independent subsets of the data
    - Split by time or wavelength
- Follow up model fitting
- Image reconstruction from multiple realisations of simulated data

Summary and perspectives

Summary and perspectives

- general ***inverse problem*** framework suitable to describe most methods;
- optimization
  - difficulties: ***non-linearity***, lots of variables (as many as pixels), ***constraints*** (non-negativity), *etc.*
  - ***direct optimization*** of the criterion is more consistent and probably more efficient
  - ***global optimization*** is required
- a priori constraints:
  - regularization: ***TV*** and ***compactness*** appear to be the most effective ($\ell_2 - \ell_1$ probably a better compromise for astronomical images)
- the future: ***multi-spectral data***
  - spectral regularization (Soulez et al. 2008)
  - much more parameters to fit, computational cost will be a big issue



(le Bouquin et al. 2009)

- other links: ***medical tomography***, ***compressive sensing***, *etc.*

Baron, F. and Young, J. S.: 2008, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 7013 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, p. 70133X

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J.: 2010, *Foundations and Trends in Machine Learning* **3**, 1

Charbonnier, P., Blanc-Féraud, L., Aubert, G., and Barlaud, M.: 1997, *IEEE Trans. Image Process.* **6(2)**, 298

Cornwell, T. J. and Wilkinson, P. N.: 1981, *Month. Not. Roy. Astron. Soc.* **196**, 1067

Donoho, D.: 2006, *Communications on Pure and Applied Mathematics* **59(7)**, 907

Fomalont, E.: 1973, *Proc. IEEE: Special issue on radio and radar astronomy* **61(9)**, 1211

Hofmann, K.-H. and Weigelt, G.: 1993, *Astron. Astrophys.* **278(1)**, 328

Högbom, J. A.: 1974, *Astron. Astrophys. Suppl.* **15**, 417

Ireland, M., Monnier, J., and Thureau, N.: 2008, in J. D. Monnier, M. Schöller, and W. C. Danchi (eds.), *Advances in Stellar Interferometry.*, Vol. 6268, pp 62681T1–62681T8, SPIE

le Besnerais, G., Lacour, S., Mugnier, L. M., Thiébaut, É., Perrin, G., and Meimon, S.: 2008, *IEEE J. Selected Topics in Signal Process.* **2(5)**, 767

le Bouquin, J.-B., Lacour, S., Renard, S., Thiébaut, E., and Merand, A.: 2009, *Astron. Astrophys.* **496**, L1

Marsh, K. A. and Richardson, J. M.: 1987, *Astron. Astrophys.* **182**, 174

Meimon, S., Mugnier, L. M., and le Besnerais, G.: 2005a, *J. Opt. Soc. America A* **22**, 2348

Meimon, S., Mugnier, L. M., and le Besnerais, G.: 2005b, *Optics Letters* **30(14)**, 1809

Narayan, R. and Nityananda, R.: 1986, *Annual Rev. Astron. Astrophys.* **24**, 127

Readhead, A. and Wilkinson, P.: 1978, *Astrophys. J.* **223(1)**, 25

Renard, S., Thiébaut, É., and Malbet, F.: 2011, *Astron. Astrophys.* **533**, A64

Rudin, L., Osher, S., and Fatemi, E.: 1992, *Physica D* **60**, 259–

Schwab, F.: 1980, in *Proc. SPIE*, Vol. 231, pp 18–25

Soulez, F., Thiébaut, É., Gressard, A., Dauphin, R., and Bongard, S.: 2008, in *16th European Signal Processing Conference*, EUSIPCO, Lausanne, Suisse

Thiébaut, É.: 2002, in J.-L. Starck and F. D. Murtagh (eds.), *Astronomical Data Analysis II*, Vol. 4847, pp 174–183, SPIE, Bellingham, Washington

Thiébaut, É.: 2008, in F. D. Markus Schöller, William C. Danchi (ed.), *Astronomical Telescopes and Instrumentation*, Vol. 7013, pp 70131I–1–70131I–12, SPIE

Thiébaut, É.: 2009, *New Astronomy Reviews* **53**, 312

Thiébaut, É. and Giovannelli, J.-F.: 2010, *IEEE Signal Process. Mag.* **27(1)**, 97