M. Juvela

Observatory, University of Helsinki

Inverse problems in the analysis of astronomical images

Inversion methods in astronomy
Helsinki University Observatory
28.2.2008

Contents

I. De-convolution: theory
- linear methods, Bayesian framework
- iterative methods
- interferometry
- Maximum Entropy
- multiscale methods

II. De-convolution: practice
- routines in reduction packages
- some examples

Convolution

- present in all observations
  - finite time resolution → convolution of time
    - detector response, temporal sampling
  - finite spectral resolution → convolution of frequencies
    - detector bandpass, channel resolution
  - point spread function → convolution of spatial coordinates
    - telescope beam, seeing, detector geometry

Convolution

- in 1D: $$(f * g)(h) = \int_{-\infty}^{+\infty} f(x)\, g(h-x)\, dx$$

  - convolution of source intensity *I(f)* with the detector response curve *R(f)*

$$I_{obs}(f) = (I_{true} * R_{det})(f) \;=\; \int_{0}^{+\infty} I_{true}(f')\;\; R_{det}(f - f')\, df'$$

  - observed intensity distribution on the sky

$$I_{obs}(\theta_{0,}\, \phi_0) \;=\; \int_0^{2\pi} d\phi \int_0^{\pi} d\theta\;\; I_{true}(\theta, \phi)\;\; B(\theta_0 - \theta, \phi_0 - \phi)$$

Convolution

- convolution can be calculated easily with Fourier transforms: **convolution theorem**

$$F(I_{obs}) = F(I_{true} * B) = F(I_{true})\ F(B)$$

... but if the filter function *B* is known, the original signal can be recovered directly

$$I_{true} = F^{-1}\left(F(I_{obs}) / F(B)\right) \quad \square$$

(this is the least squares solution)

De-convolution

- How would this work in practice
  - Example 1: deconvolution of a 1 D signal
  - Example 2: deconvolution of a 2D map
  - the noise should be part of the model

$$F(I_{obs}(x,y)) = F(I_{true})\ F(B) + F(n(x,y))$$

- least squares should be fine for gaussian noise

De-convolution

- conclusion:
  - de-convolution is trivial – in the sense of getting a function whose convolution is equal to the observation
  - result is extremely sensitive to the noise (and the knowledge of the convolving function)
    - solution also tries to fit all the noise
  - deconvolution becomes easily an ill-posed problem => need some regularization

Convolution

- usually one works with discrete quantities

$$D_i = \sum_j H_{i,j} I_j + \epsilon_j$$

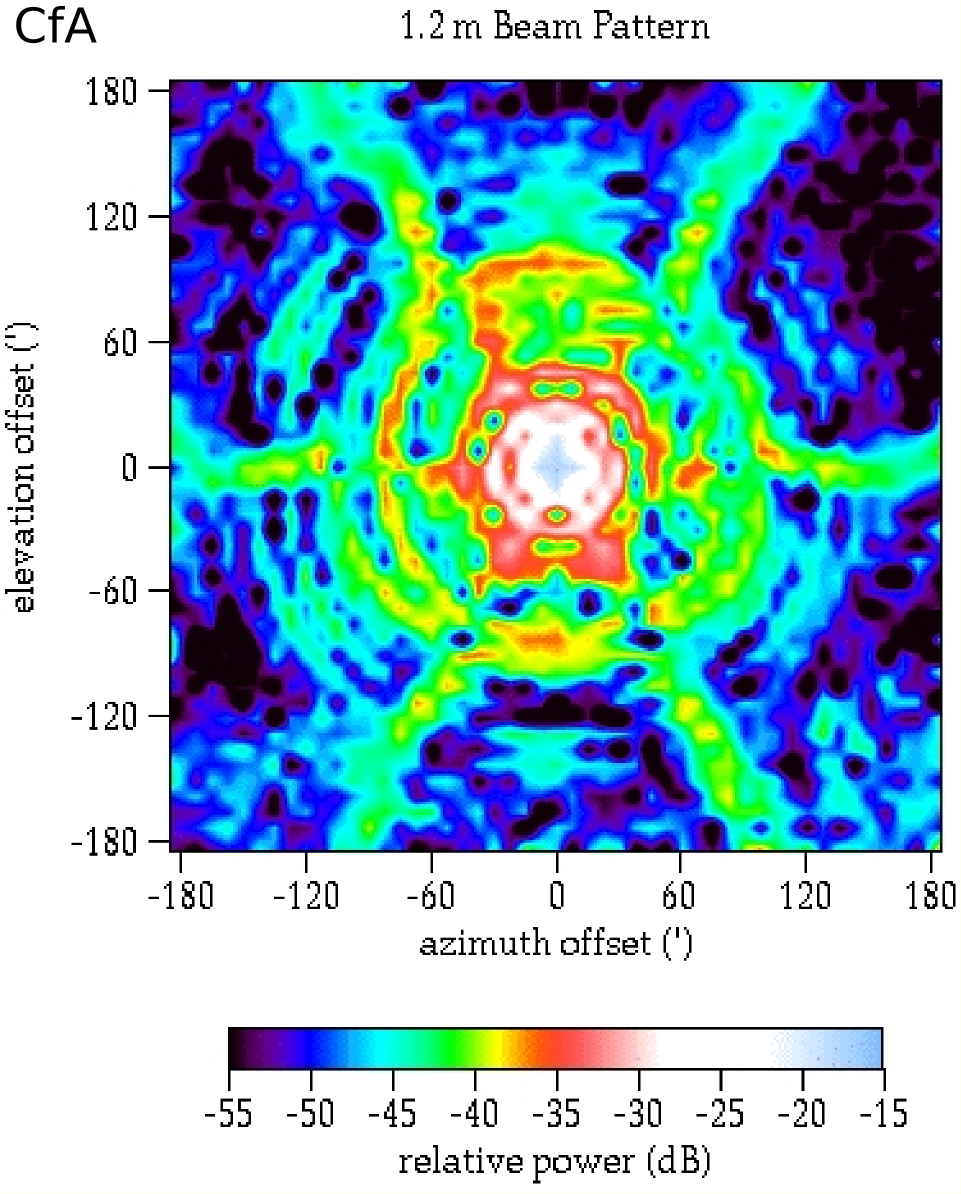
  - *D* is the observed image, *H* elements of the point spread matrix, *I* the true signal, and ε the noise in the observed data elements

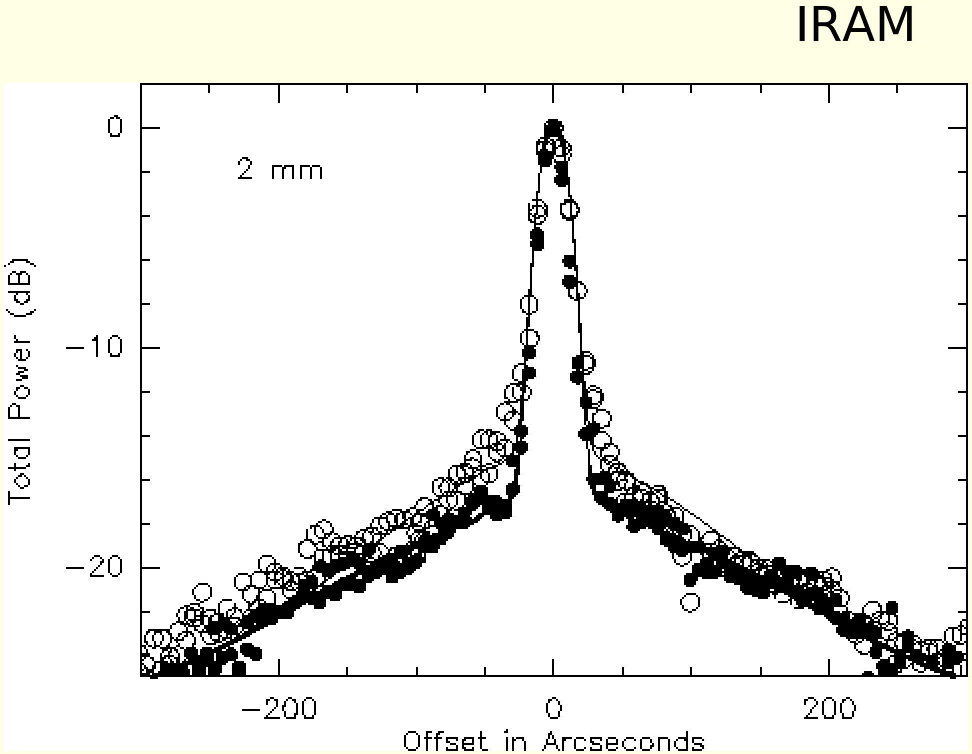
Naive solution

- naively one might try a least squares solution (assuming the errors were normally distributed)

$$min_I \frac{\|H * I - D\|^2}{\sigma^2}$$

- as the previous examples showed, this does not usually work
  - oscillatory solutions
  - might be acceptable if number of data points > number of points in the reconstructed image *and* the noise level is small


CfA
1.2 m Beam Pattern
elevation offset (')
azimuth offset (')
180
120
60
0
-60
-120
-180
-180 -120 -60 0 60 120 180
-55 -50 -45 -40 -35 -30 -25 -20 -15
relative power (dB)


IRAM
2 mm
Total Power (dB)
0
−10
−20
−200
0
200
Offset in Arcseconds

Smoothed least squares

- in the basic regularization we require the result image to be 'smooth'
  - no rigorous definition of smoothness
- the solution is found by minimizing

$$min_I \ [\frac{\|H * I - D\|^2}{\sigma^2} \ + \ \lambda \phi(I)]$$

  - λ is a smoothing parameter
  - ϕ is smoothing function, for example

$$\phi(I) = \sum_j (I_j - I_{j+1})^2$$

Smoothed least squares

- in Tikhonov regularization one minimizes function

$$J(I) = \left\| D(x,y) - (H*I)(x,y) \right\| + \lambda \left\| P*I \right\|$$

  - first term is χ^2
  - second term contains convolution of the result image with a high pass filter
  - λ is regularization parameter that determines balance between exact fit and image smoothness
- the solution (*note syntax:* $F(x) = \tilde{x}$)

$$\tilde{I} = \frac{\tilde{H}^* \tilde{D}}{|\tilde{H}|^2 + \lambda |\tilde{P}|^2}$$

Smoothed least squares

- more generally one might write the smoothing function as a quadratic form

$$\phi(f) = f^T C f$$

- the solution of the resulting equation

$$min_I \; [\frac{\|H I - D\|^2}{\sigma^2} \; + \; \lambda I^T C I]$$

becomes $I = (H^T H + \lambda C)^{-1} H^T D = K(\lambda) D$

  - solution is found quickly with linear algebra
  - least squares solution recovered with λ=0
  - stable solution as soon as λ 'sufficiently' large

Smoothed least squares

- the main problem is the selection of the smoothing parameter
  - often an ad hoc value - what looks right
  - below we follow Thompson & Graig -92
- objectively λ could be chosen so that correct χ^2-value is recovered $\|\hat{g}-g\|^2=n\sigma^2$
  - g the observed data, $\hat{g}$ reconstructed image
  - in practice this leads to too smooth solutions
- different criteria developed for finding the optimal smoothing

Smoothing parameters: EDF

- the reason the normal χ^2 criterion fails is that counts d.o.f. of the fit but ignores the d.o.f. of the observed image

$$\|\hat{g}-g\|^2+\langle|\hat{g}-\langle\hat{g}\rangle|^2\rangle=n\sigma^2 \qquad (1)$$

  - χ^2 criterion ignores the second term => fit becomes less precise => resulting image is too smooth

- the variance can be estimated if the probability distribution of the recovered image is known
  - see Thompson & Graig (1992)

Smoothing parameters: EDF

- when *I=K(λ) D*, the variance term becomes

$$\langle |\hat{g} - \langle \hat{g} \rangle|^2 \rangle \;=\; \sigma^2 tr(K(\lambda)) \;=\; \sigma^2 tr(H(H^T H + \lambda C)^{-1} H^T)$$

  - λ is selected so that Eq. 1 holds

$$\|H\hat{f} - g\|^2 = \sigma^2 tr(I - K(\lambda))$$

  - result can be generalized for non-quadratic smoothing functions (Thompson & Kay -92)

Smoothing parameters: CB

- Craig & Brown (1990) use the same criterion $\|\hat{g}-g\|^2+\langle|\hat{g}-\langle\hat{g}\rangle|^2\rangle=n\sigma^2$ but calculate the second term according to the stability of the solution
- the probability of a realization g_k is

$$P(g_k)\propto\exp\left(\frac{1}{2}\left\|\frac{Hf-g_k}{\sigma}\right\|^2\right)=\exp\left(\frac{1}{2}\left\|\frac{g-g_k}{\sigma}\right\|^2\right)$$

  - the latter form follows when *Hf* is replaced with our best estimate, i.e., the actually observed map

Smoothing parameters: CB

- the previous leads to the result

$$\langle |\hat{g} - \langle \hat{g} \rangle|^2 \rangle = \langle |\hat{g} - \langle \hat{g}_k \rangle|^2 \rangle$$

$$= \int (\hat{g}_k - \langle \hat{g}_k \rangle)^T (\hat{g}_k - \langle \hat{g}_k \rangle) \exp(-\frac{1}{2} \left(\frac{g_k - g}{\sigma} \right)^2) d(g_k)$$

$$= \sigma^2 tr(K(\lambda) K(\lambda))$$

  – smoothin parameter is found based on

$$\|H \hat{f} - g\|^2 = \sigma^2 tr(I - K(\lambda) K(\lambda))$$

  – larger filter factor than EDF, asymptotical difference smaller than a factor of two

Smoothing parameters: BAS

- Bayesian derivation by Gull (1988)
  - several formulations (one method identical with EDF)
  - one particular form

$$\|H\hat{f}-g\|^2+\lambda\hat{f}^T C\hat{f}=n\sigma$$

    - *H* was the beam, *C* the constraints
    - this can be evaluated without calculating the trace of matrix *K* => suitable for large problems

Smoothing parameters

- conclusions of Thompson & Graig
  - easy problem = low blur
    - EDF reconstruction close to the best possible
    - CB oversmooths slightly (some bias)
    - BAS worse, consistently undersmoothed
  - harder problem = large blur
    - EDF sensitive to σ, sometimes *grossly* undersmoothed
    - CB appears slightly more robust
    - BAS undersmoothed (features produced by noise) but insensitive to the value of σ
      - the χ^2 method produces grossly oversmoothed results

Linear regularization methods

- some problems
  - Gibbs oscillations near discontinuities
  - hard to use any a priori information (even positivity constraints)
  - regularization usually through smoothing: leads to loss of resolution

Some regularization schemes

$$\left\|C\,I\right\| = \sum\sum I(x,y) - \frac{I(x-1,y)+I(x+1,y)+I(x,y-1)+I(x,y+1)}{4}$$

- sometimes called (simultaneous) autoregressive model
- equivalent to prior $\propto \exp\left(-\frac{\alpha}{2}\left\|C\,I\right\|\right)$

– examples of other forms

- Charbonnier et al. (1997)
- Moline et al. (1996, 2001, 2000)

Bayesian framework

– the Bayes formula of probabilities

$$P(I|D)=\frac{P(I)P(D|I)}{P(D)}$$

– in this case we interpret *D* as the observed data and *I* as the true intensity

– solution found by maximizing posterior probability *P(I|D)*

- *P(I)* is prior probability of given solution *I*
- *P(D|I)* is the probability of data when model *I* is given (by itself would often lead to χ^2 minimization)
- *P(D)* is merely a normalization factor

Bayesian framework

– the maximum likelihood estimate would be

$$ML(I) = max_I \;\; p(D|I)$$

while the solution from Bayes formula is

$$ML(I) = max_I \;\; p(D|I)\, p(I)$$

- MAP = maximum a posteriori solution
- denominator *P(D)* does not affect the solution
- ML = MAP with constant prior

Bayes: Gaussian noise

– for gaussian (=normally distributed) noise

$$p(D|I) = (\ldots)\exp\left[-\frac{1}{2}\left(\frac{D - I \; * \; H}{\sigma}\right)^2\right]$$

- in unconstrained case this leads to χ^2 minimization

$$max\ p(D|I) = max \exp\left(\frac{-1}{2}\chi^2\right) \Rightarrow min\,\chi^2$$

- in constrained case regularization can be built in the iterative optimization algorithm
  - Landweber / successive approximations / Jacobi method
  - number of iterations ~ smoothness of solution
  - include other constraints, e.g., positivity

Bayes: Gaussian noise

– in the special case that both object and noise are normally distributed with zero mean, the Bayes solution leads to Wiener filtering

$$\tilde{I} = \frac{\tilde{H}^* \tilde{D}}{|\tilde{H}|^2 + \sigma_N^2 / \sigma_o^2}$$

– as before, *H* is the beam, *D* the observations, and *I* the recovered intensity

– σ_N^2 the noise variance and σ_o^2 the object variance

- Wiener filtering is very fast to calculate – and optimal in case of stationary, Gaussian signal
- ... but causes artifacts (e.g., rings around point sources) and needs noise estimates in the *frequency* space – $\sigma = \sigma(\nu)$!

Bayes: Poisson noise

– Poisson distribution is $$p(k)=\frac{\mu^k}{k!}e^{-\mu}$$

- *k* is the discrete number of events
- μ is the expectation value (and variance)

and the probability becomes

$$p(D|I)=\prod_{x,y}\frac{\left[(H*I)(x,y)\right]^{D(x,y)}e^{-(H*I)(x,y)}}{D(x,y)!}$$

– ML estimate is found by setting derivate of ln *p(D|I)* to zero

$$\frac{\partial \ln p(D|I)}{\partial I}=0 \quad \Rightarrow \quad \frac{D}{H*I}*H^{*}=1$$

Bayes: Poisson noise

– this leads to an iterative algorithm

- multiply both sides with *I*, evaluate left side using an old estimate on the right hand side

$$I^{(n+1)} = I^{(n)} \left(\frac{D}{H * I^{(n)}} * H^{*} \right)$$

- this is the famous **Richardson-Lucy** algorithm (Lucy 1974) or the **expectation maximization** (EM) method
- the flux is conserved and the solution is always non-negative
- still only a maximum likelihood solution

Bayes: Poisson noise

– if we denote with *M* the true solution, the probability of given model *I* becomes

$$p(I) = \frac{\prod M^I e^{-M}}{I\,!}$$

and the MAP solution is

$$I = M \exp\left\{\left[\frac{D}{H * I} - 1\right] * H^{*}\right\}$$

setting $M = I^{(n)}$ one obtains an iterative formula

$$I^{(n+1)} = I^{(n)} \exp\left\{\left[\frac{D}{H * I^{(n)}} - 1\right] * H^{*}\right\}$$

Iterative regularized methods

- let *I* be the recovered intensity
- constraints can be presented as a function P_C so that we require, e.g.,
  - positivity of solution
  - object belongs to spatial domain *D*

$$P_C(I(x,y)) = \begin{cases} I(x,y), & \text{if } (x,y) \in D \\ 0, & \text{otherwise} \end{cases}$$

  - solution is band limited

$$P_C(\tilde{I}(x,y)) = \begin{cases} \tilde{I}(x,y), & \text{if } \nu < \nu_0 \\ 0, & \text{otherwise} \end{cases}$$

Iterative regularized methods

- the constraints can be implemented easily in iterative schemes
  - smaller number of iterations acts also like regularization
- below are examples of regularized versions of the following algorithms
  - Jansson-Van Cittert
  - Landweber
  - Tikhonov
  - Richardson-Lucy

Iterative regularized methods

- the van Cittert iteration is very simply

$$I^{(n+1)} = I^{(n)} + \alpha\left(D - H * I^{(n)}\right)$$

  - convergence parameter can be set $\alpha \sim 1$
  - may converge in a few iterations, diverges in the presence of noise

- Jansson (-70) considered constraint $A < I < B$

$$\alpha \to C\left[1 - 2\frac{|I^{(n)} - (A+B)/2|}{B-A}\right]$$

  - more generally $I^{(n+1)} = P_C\left\{I^{(n)} + \alpha\left(D - H * I^{(n)}\right)\right\}$

Iterative regularized methods

- regularized Landweber iteration

$$I^{(n+1)} = P_C\left\{ I^{(n)} + \gamma H^{*} * \left(D - H * I^{(n)} \right) \right\}$$

  - ~steepest descent minimization (Jacobi method)
  - $H^{*}(x,y) = H(-x,-y)$

- regularized Richardson-Lucy

$$I^{(n+1)} = P_C\left\{ I^{(n)} \left[\frac{D}{H * I^{(n)}} * H^{*} \right] \right\}$$

Iterative regularized methods

- Tikhonov solution is obtained using gradient function

$$\nabla J = [H^* * H + \mu P^* * P] * I - H^* * D$$

  - P is a high pass filter

– normal iteration is

$$I^{(n+1)} = I^{(n)} - \gamma \nabla J$$

and the regularized version, not surprisingly,

$$I^{(n+1)} = P_C \left\{ I^{(n)} - \gamma \nabla J \right\}$$

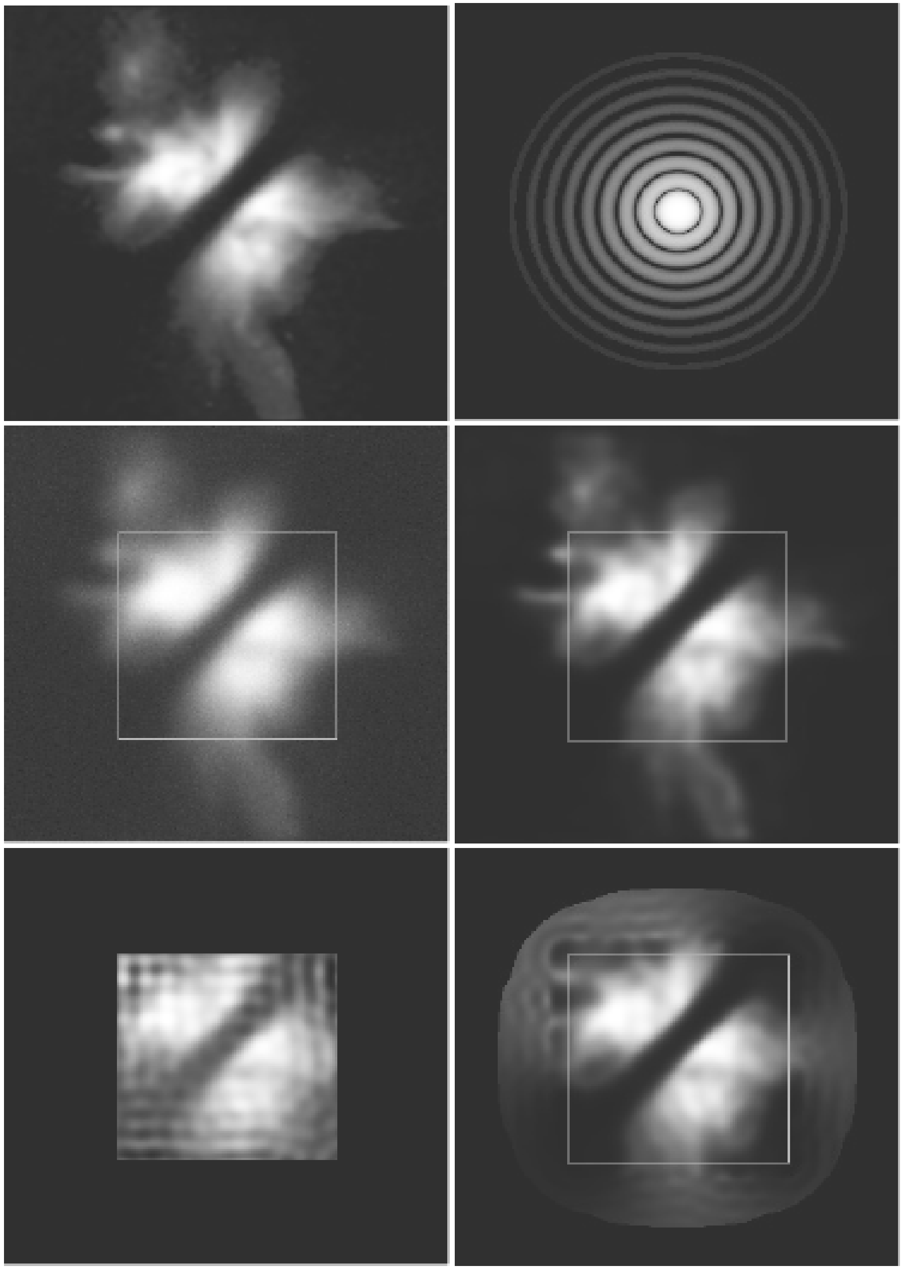
Edge effects

- FFT based methods are computationally efficient in de-convolution
  - Richardson-Lucy requires 4 FFT transformations and total cost is $O(N^2 \log N)$
- in the case of extended emission edges cause ripple in the solution
  - artifacts can be decreased by ×10 by using reflective or anti-reflective boundary conditions
    - changes the result – what did the real beam see outside the image?

Edge effects

- the boundary pixels depend on the unknown intensity outside the FOV
- Bertero & Boccacci (2005):
  - RL is used to deconvolve a larger image, which improves the reconstruction inside FOV
  - brightness values outside measured map as free parameters
  - fast implementation using FFT:s

Edge effects

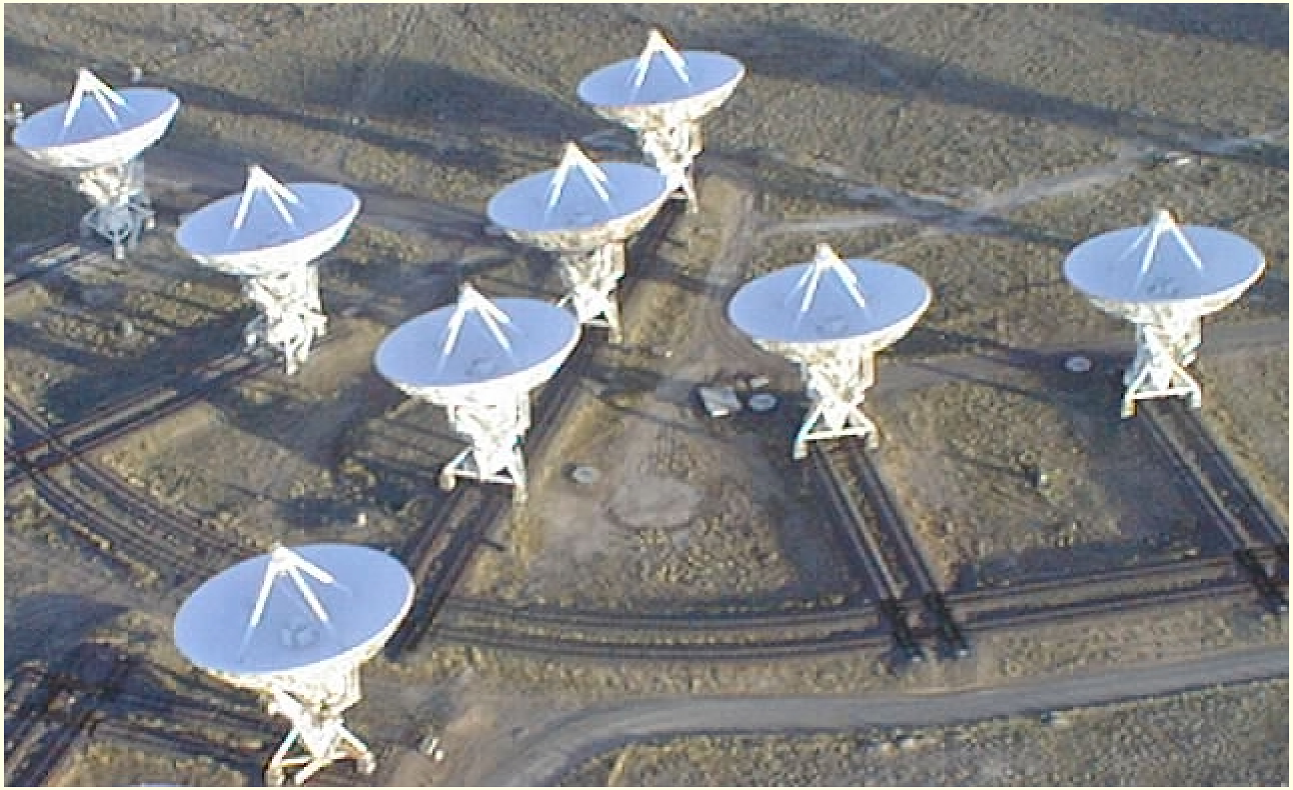


original image and the beam

convolved image with Poisson noise and RL reconstruction of full image

reconstruction of smaller area with RL and with the method of Bertero & Boccacci

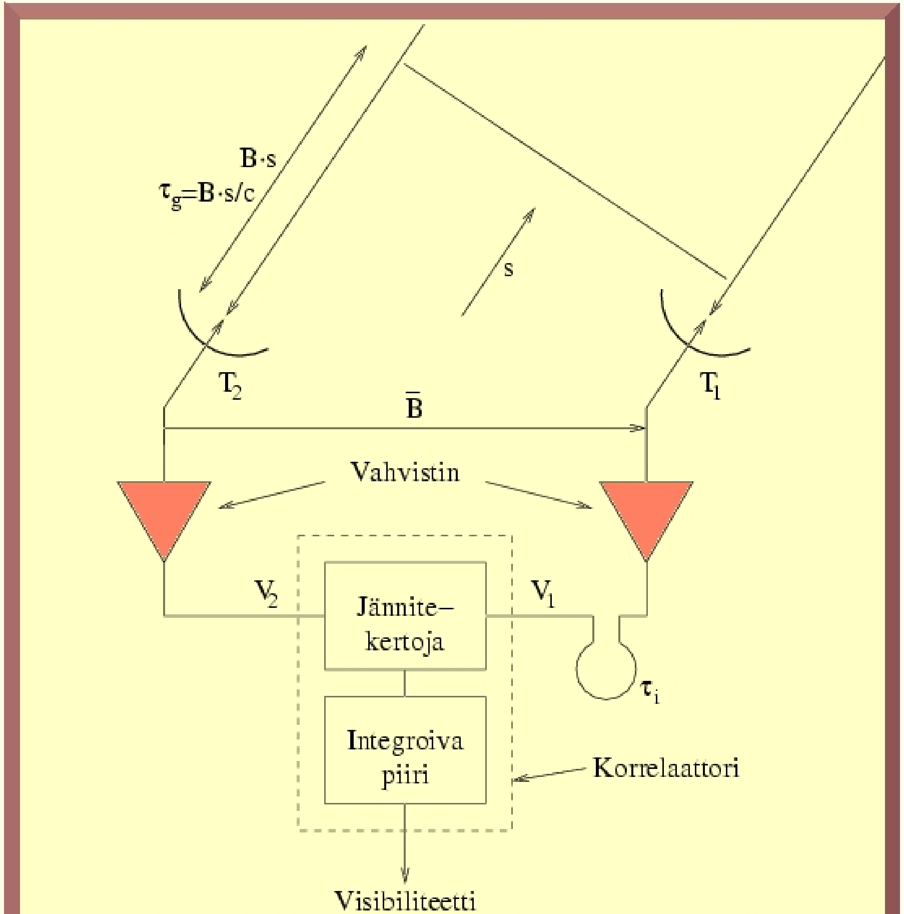
Interlude: Interferometry



VLA/NRAO

Interferometry

- correlation of signals observed by two or more antennas



Kuva 6.1: Korreloiva interferometri.

- same wavefront is observed in antennas with a phase shift

$$\tau_g = \frac{1}{c}\vec{B}\cdot\vec{s}$$

- the correlator calculates time average of the product of measured voltages *V*
- V^2 is proportional to the power of the radiation
- for a point source the cross correlation is

$$R_{xy} \propto \sqrt{w_1 w_2} e^{i 2\pi \nu \tau}$$

Interferometry

- the phase shift is set to zero at a phase center close to the object
- the corrected cross correlation is called the *visibility*
  - this depends on the source intensity and the angles between the source direction and the baseline *B* connecting the antennas
- usually coordinates *u* and *v* are used
    - *u* = east-west length of *B* as seen from the phase centre
    - *v* = corresponding length of the north-south projection

Interferometry

- in aperture synthesis the observed visibilities are used to derive a continous map of the sky around the phase centre
- for an intensity distribution the visibility is

$$V(u,v) = \int I_\nu(x,y) A_e(x,y) e^{i2\pi(ux+yv)} dx\, dy$$

  - *I(x,y)* is the true surface brightness
  - *A*(x,y) is the effective collecting area of the telescopes, *including their beam pattern*
  - ... one observes a Fourier transform of the surface brightness !

Interferometry

- when we look at a small area surrounding the phase centre, the beam of individual telescopes is almost constant

$$V(u,v)/A_e = \int I_\nu(x,y) e^{i2\pi(ux+yv)} dx\, dy$$

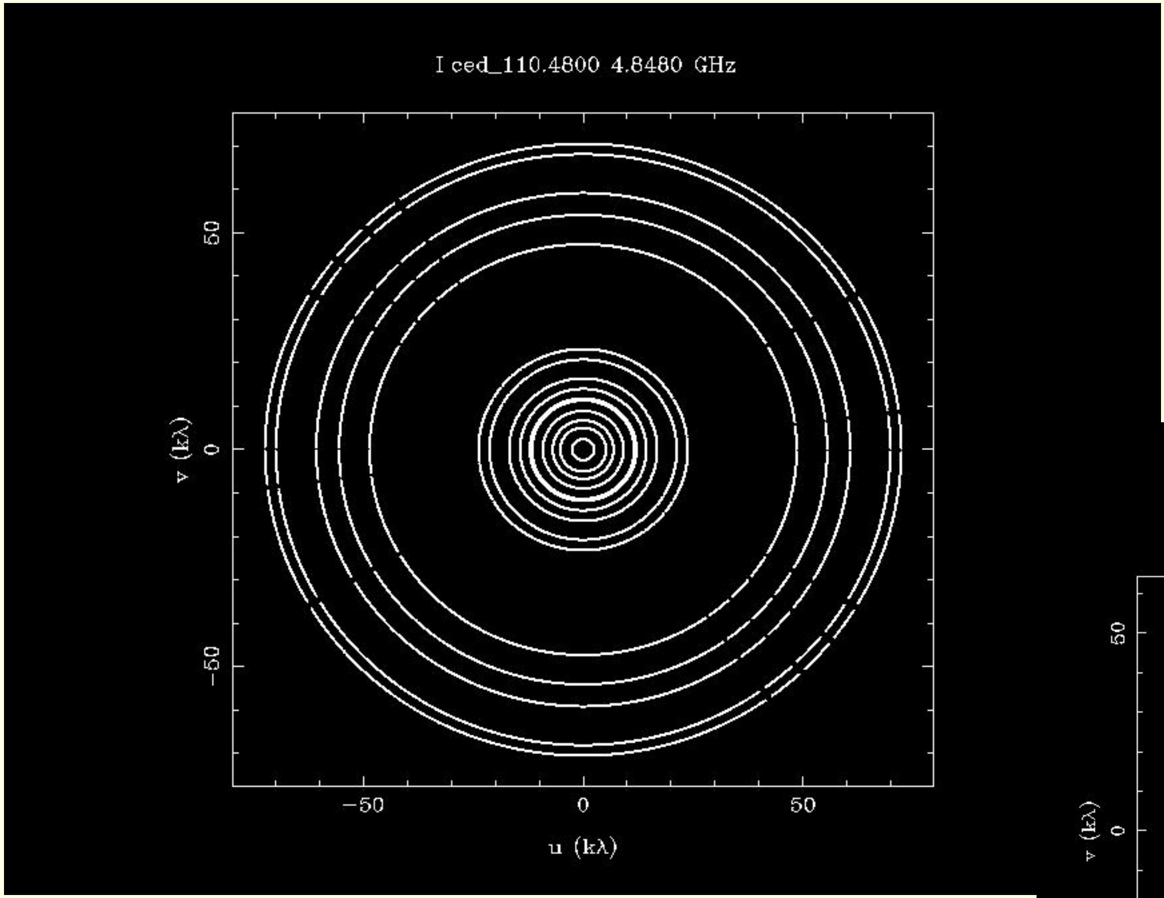
- if visibilities were observed for all *u* and *v*, the intensity would be obtained by a direct Fourier transform

$$I_\nu(x,y) = \int\int V(u,v)/A_e \; e^{-i2\pi(ux+vy)} du\, dv$$

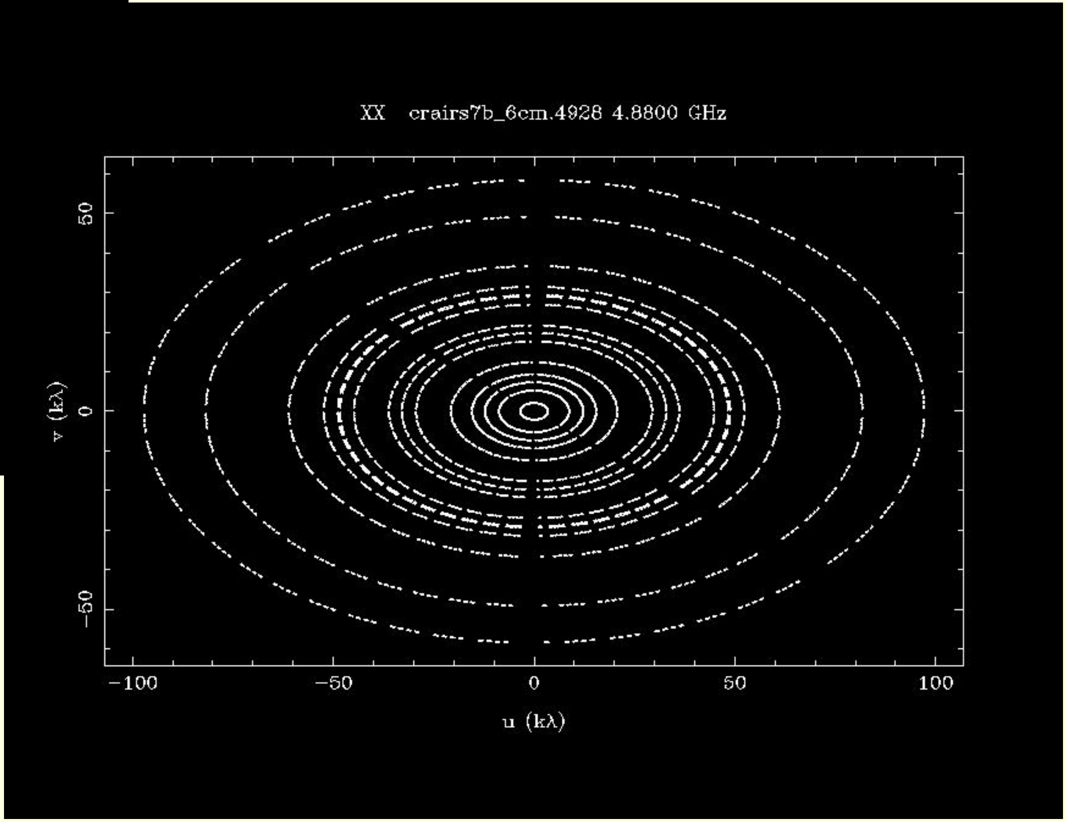
Interferometry

- in practice only a small part of the (u,v) plane is covered
  - each antenna pair gives instantaneously only one position (u,v)
    - ... and the corresponding point (-u,-v)
    - because visibilities are real, $V(-u,-v)=V^{*}(u,v)$
  - as the Earth rotates, each antenna pair draws a curve in the (u,v)-plane

Interferometry



(u,v)-coverage in two cases



Interferometry

- the measured visibility can be seen as a convolution between true visibility and a mask g
  - g=1 where we have measurements, g=0 elsewhere
  - by definition the product $g(u,v)V(u,v)$ is known everywhere and the Fourier transforms can be performed

Interferometry

- according to convolution theorem

$$F(g(u,v)V(u,v)) = Con(P_{syn}(l,m),\ \ M(l,m))$$

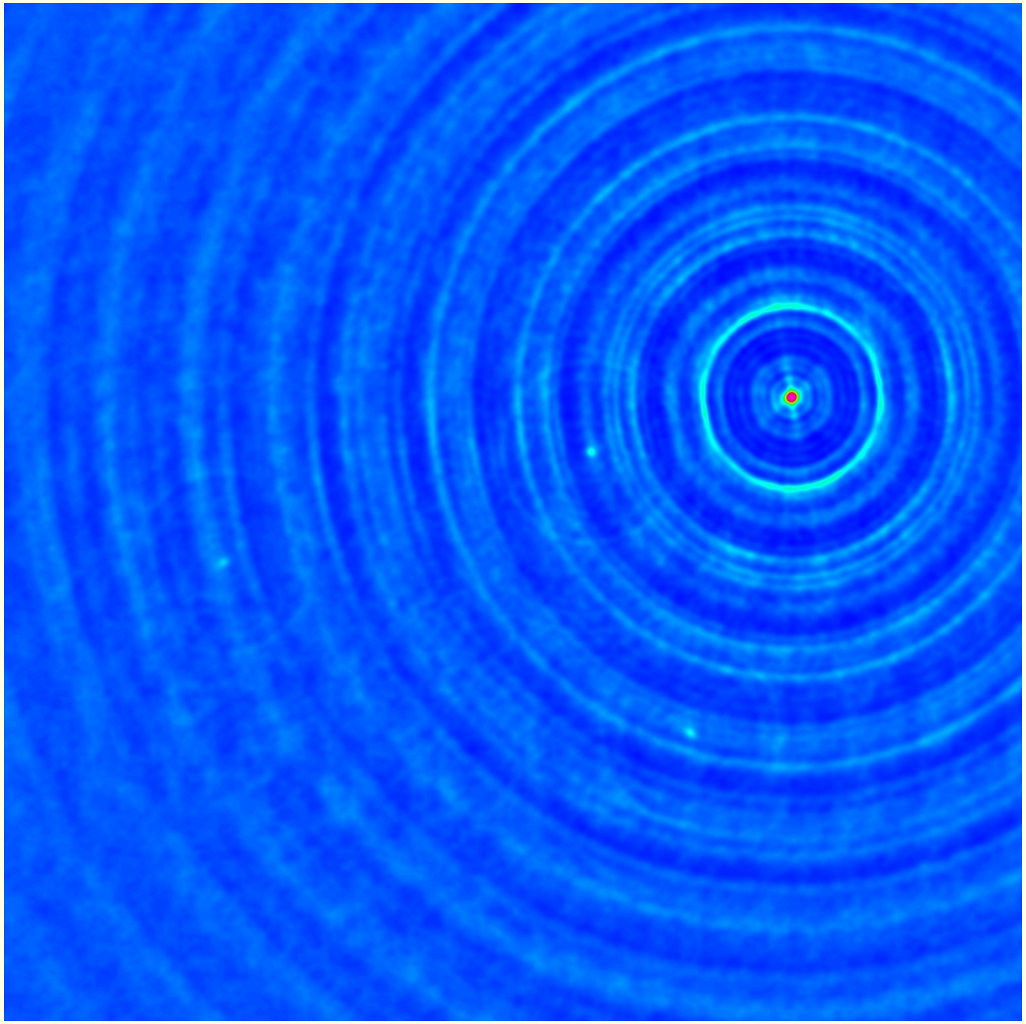
  - P_{syn} is the Fourier transform of *g* = **dirty beam**
  - this defines the final resolution of the recovered map
  - *M* is essentially the product of the true intensity and the beam pattern of an individual antenna
    - *g* is also called the *grading function* of the synthesized aperture

Interferometry

- the recovered 'dirty image' contains artifacts caused by the incomplete sampling of the (u,v) plane
  - each point source produces an image of the dirty beam
  - one must somehow fill in the missing visibilities (regularization) or *clean* the final image

Interferometry

- dirty image:



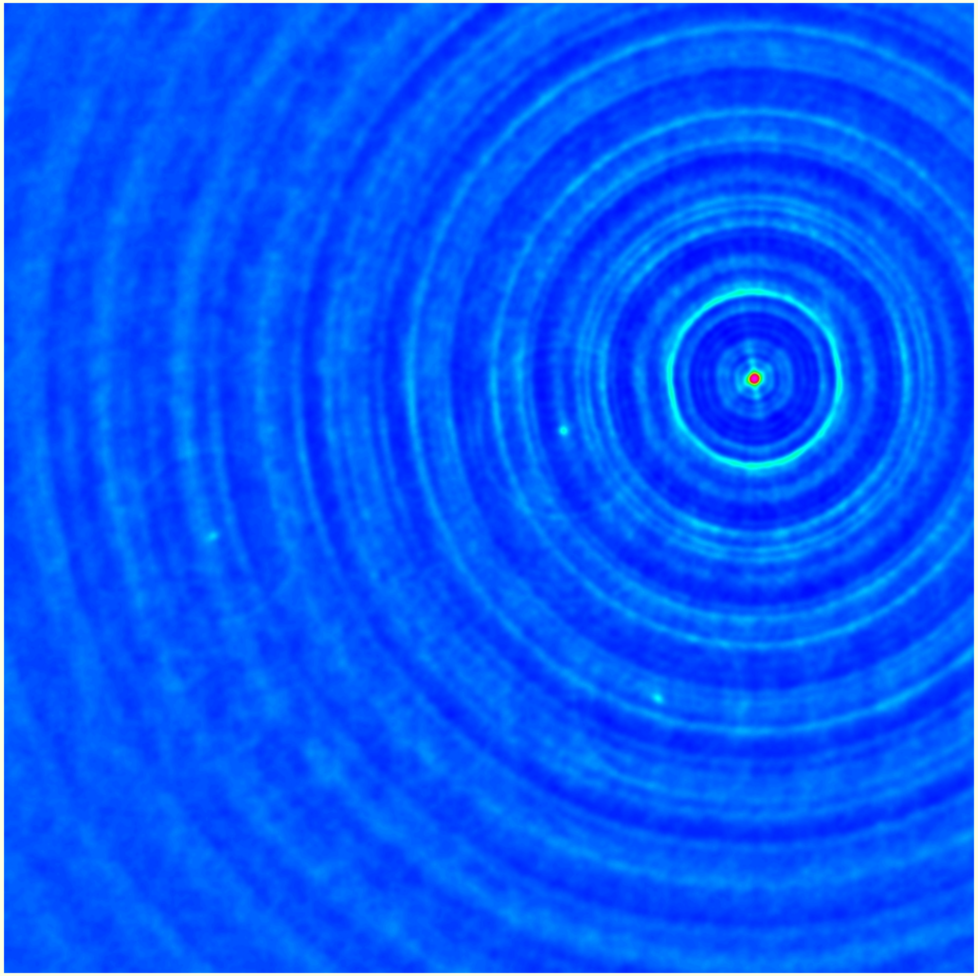
CLEAN

- the CLEAN algorithm (Högbom 1974) models image as a sum of point sources
    1. find an intensity peak in the (dirty) image
    2. subtract the peak = (dirty) beam multiplied with a damping factor
    3. repeat there are no more peaks above specified level
    4. convolve point source model with idealized 'CLEAN beam' (e.g., central lobe of (dirty) beam)
        – above 'dirty image' and 'dirty beam' refer to interferometric observations

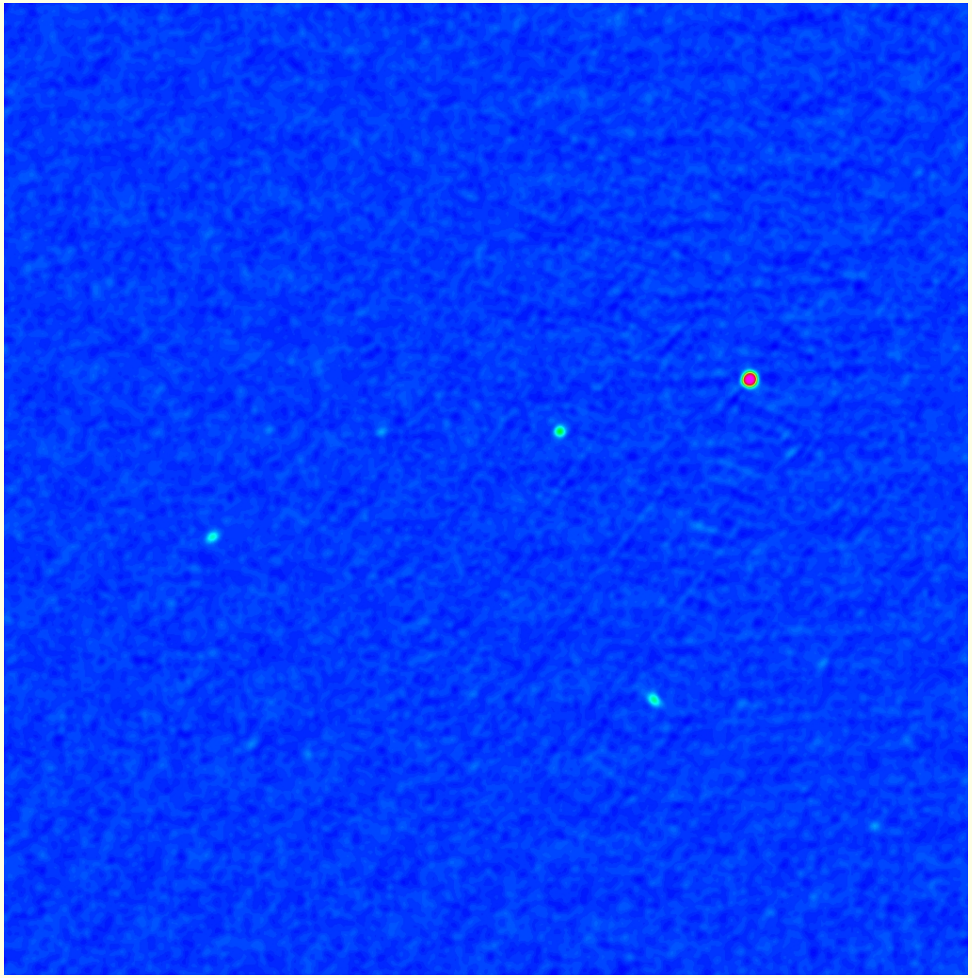
CLEAN

- the residual image may be added to CLEAN image (as a check)
- instead of the real space image one can work in Fourier space using FFTs (Cark 1980) or, in the case of interferometry, directly with visibilities
- solution is unique
  - ... if there is no noise and if the number of visibility measurements is larger than the number of image elements
  - in principle superresolution should be possible
  - in practice (because of the noise and other factors) CLEAN performs badly in this respect

dirty image



cleaned and restored image



CLEAN

- some negative aspects
  - slow for large images
  - no handle on the image statistics
  - spurious peaks caused by the noise
  - depression in surface brightness around strong sources
  - fluctuations in extended emission
  - final convolution with 'CLEAN beam' is basically an ad hoc procedure to produce nice images
    - this affects also the relative scaling between CLEAN image and residuals!
  - if there is a good source model for extended emission, that should be subtracted before CLEANing the image

Maximum entropy

- entropy is a measure of disorder
- in image reconstruction one would like to obtain an image that contains all information available from the observations – but no additional structure
- maximation of image entropy should guarantee this
  - 'maximally noncommittal'
  - 'as featureless as the data allow'

Maximum entropy

- the problem is still the inversion of eq.

$$D_i = \sum_j H_{i,j} I_i + N_i$$

  - *D* is the observed quantity, in images usually the intensity
  - *H* is the point spread function
  - *N* (additive) noise
- the entropy of the recovered image is

$$S = \int \int f[I(x,y)]\, dx\, dy \qquad S = \sum_{i,j} f[I_{i,j}]$$

  - *f* is some function (not quite a unique definition!)

Maximum entropy

- entropy is related to the probability of a state
  - entropy measures the number of ways a given state can be realize
  - entropy is additive while probabilities are combined multiplicatively => entropy should be proportional to the logarithm of probability
    - if there are W ways to realize a state, each with probability $p_w = 1/W$, entropy is $S = \ln W = -\ln p_w$
    - when alternatives have probabilities p_i the average becomes

$$S = -\langle \ln p_i \rangle = -\sum_i p_i \ln p_i$$

Maximum entropy

- if original variable *X* is transformed to *Y*

$$S=-\int p(X)\ln[p(X)]dx=-\int q(Y)\ln[q(Y)/J(Y)]dY$$

  - *J* is the Jacobian of the transformation, $J=dX/dY$
- this suggests that entropy should be written in form

$$S=-\int p(X)\ln[p(X)/p_0(X)]dX$$

  - $p_0(X)$ is 'prior', which is analogous with the state degeneracy *g* in the discrete case

$$S=-\sum_i p_i \ln(p_i/g_i)$$

Maximum entropy

- the prior makes entropy independent of the scaling
- when prior is constant (flat image), entropy is maximum when p_i is constant
  - more generally, maximum corresponds to state where p is proportional to its prior
- in practice different forms of entropy are used

$$f(I) = -\ln I$$

$$f(I) = -I \ln I$$

Bayesian derivation

- the Bayes equation of probabilities says

$$P(I|D)=\frac{P(I)\,P(D|I)}{P(D)}$$

  - *P(I)* is the prior probability $P(I)\propto\exp[S(I)]$
  - *P(D|I)* is the probability of observations when image *I* is given; in case of gaussian white noise

$$P(D|I)\propto\prod_r \exp\left[-\frac{1}{2}\sum_i\left(\frac{H_{ri}I_i-D_r}{\sigma}\right)^2\right]$$

    - note that the denominator *P(D)* does not depend on the solution and is merely a normalization factor

Bayesian derivation

- when we take a logarithm we obtain probability

$$\ln P(I|D) = S(I_i) - \frac{1}{2}\sum_r \left(\frac{\sum_i H_{ri} I_i - D_r}{\sigma_r} \right)^2$$

  - this is a sum of entropy and the χ^2 value
  - our solution would correspond to the maximum of this probability
  - conversely, we could minimize $\chi^2 - S$
  - note: assumes uncorrelated errors and normal error distribution!

Entropy functions

- in the framework of power spectrum estimation Burg (1967) came up with the formula corresponding to $S_1 = \int \ln I$
  - there *I* corresponded to terms of power spectrum that can be seen as independent Gaussian random variables
  - probability is product of normal probabilities and the entropy is proportional to logarithm of variance
  - same form can be derived for images in the limit of high photon numbers

Entropy functions

- consider an image with pixel intensities I_i
- with total flux of the image D_0=sum(I_i), we can define fractional intensity $f_i = I_i / D_0$
  - the image can be constructed in $W=exp(S)$ different ways, which leads to

$$S_2 \approx -N \sum_i f_i \ln f_i = -(N/D_0) \sum_i \ln I_i (+constant)$$

  - the same formula can also be derived from thermodynamic entropy in the limit of low photon numbers
    - valid for high frequencies while S_1 might be more appropriate at radio frequencies?

Entropy functions

- the two formulations are not necessarily contradictory
  - correspond to different probability distributions one can attribute to the same image ?
  - both go to infinity when intensity goes to zero: forces results to be positive
    - there are generalizations for cases with negative intensities
  - both have negative second derivative which suppresses rapid variations

Entropy functions

– still another formula (Gull & Skilling 1991)

$$S = \sum_x \sum_y \left\{ I(x,y) - M(x,y) - I(x,y) \ln \frac{I}{M} \right\}$$

- M prior image, I our solution
- entropy reaches maximum (value zero) when image is equal to the prior

ME-approach to interferometry

- intensity *I(x,y)* is to be reconstructed based on measured function *F(D)(u,v)*
  - *D* is in real space and *F(D)* is its Fourier transform in the *(u,v)*-space so that *F(I)(u,v)=F(D)(u,v)* for all *measured u* and *v*
  - **ignoring the noise** we can maximize

$$\int\int f(I)\,dx\,dy + \sum_{u,v} \lambda_{u,v} \left(\int\int I \exp[-i2\pi(ux+vy)]\,dx\,dy - \tilde{D}(u,v) \right)$$

    - Lagrange multipliers λ enforce the constraints
  - derivation wrt *I(x,y)* gives

$$f'[I(x,y)] = -\sum_{u,v} \lambda(u,v) \exp[-i2\pi(ux+vy)] \equiv J(x,y)$$

ME-approach

- the ME image is found by formally solving this equation

$$I(x,y) = f'^{-1}[J(x,y)] \equiv g[J(x,y)]$$

  - depending on the selected form of entropy we have either

$$S = \ln I \;\Rightarrow\; S' = 1/I \;\Rightarrow\; g(J) = 1/J$$

or

$$S = -I \ln I \;\Rightarrow\; S' = -\ln I + 1 \;\Rightarrow\; g(J) = \exp(-1-J)$$

ME-approach

- the result has the following properties
  - in J(x,y) are only Fourier components corresponding to measured (u,v)
    - in the restored image missing values must be created by the non-linearity of the *f* '
    - the degree of the non-linearity depends on absolute value of *I*
  - if we add an offset in the intensities, the recovered map will be different
  - the slope of *f* ' is small at large values of *I* => peaks will be narrow ('superresolution')
  - the slope is large at small *I* => small scale variations are suppressed

ME-approach

– around extreme points Taylor expansion gives for *f=-ln I*

$$I(x,y)=\frac{1}{a(x-x_0)^2+2\mathrm{b}(x-x_0)(y-y_0)+c(y-y_0)^2+d}$$

and for *f = - I ln I*

$$I(x,y)=\exp[-a(x-x_0)^2-2\mathrm{b}(x-x_0)(y-y_0)-c(y-y_0)^2-(d+1)]$$

– in other words, in details the first definition of entropy leads to **Loretzian** peaks, the latter to **Gaussian** peaks

ME-approach

- the resolution depends on the intensity level: it is largest for the highest peaks
  - peak height is less reliable than the values of integrated flux
- suppression of small scale variation is best at low intensities
  - the sidelobes of a point source are not well suppressed if the source is located on an elevated plateau
- there can be spurious peaks around absorption features

ME-approach, noisy data

- in ME analysis one can resort to the old scheme: smooth until one gets expected value of $\chi^2 = \Omega \approx$ number of independent data points
  - the constraint $$\chi^2 = \sum_{u,v} \left[\frac{\tilde{I}(u,v) - \tilde{D}(u,v)}{\sigma(u,v)} \right]^2 = \Omega$$

    is again forced with the help of Lagrange multipliers when maximizing

    $$\int \int f(I)\, dx\, dy \;-\; \lambda(\chi^2 - \Omega)$$

    - 'least squares MEM'

ME-approach, noisy data

- the function *J* is still band limited
  - fourier coefficients of *F(J)* are zero except for the measured *(u,v)* coordinates
- the model predictions *differ* from the data

$$\tilde{J}(u,v) - \lambda \frac{\tilde{I}(u,v) - \tilde{D}(u,v)}{\tilde{\sigma}^2(u,v)} = 0$$

  - in the fourier space the residuals *F(I)-F(D)* are not random
    - highly correlated, negative at peaks and in regions of low intensity mostly positive, smoothing out variations of *I*

ME-approach, noisy data

- flux is transferred from peaks to background
  - the relative effect is largest for small peaks
    - at large S/N the residuals are larger than for the original data by a factor of sqrt(2) !
  - the solution is always biased towards the prior (usually flat distribution)
- additional constraints can be added for flux conservation of the whole image
- resolution depends on the S/N ratio
  - MEM is capable of superresolution
    - but how to tell, whether all the structures are real?
  - final MEM image can be convolved with a gaussian to get a more uniform resolution

ME-approach, noisy data

- comparison with CLEAN
  - MEM is biased, CLEAN is not; bias results from the fact that data (including he noise) is not modelled exactly
  - bias is acceptable as a trade-off with smaller variance and can be decreased by convolving the original data
  - MEM minimizes pixel variance so that the images are smoother than CLEAN images
  - CLEAN is poor for extended emission, MEM has problems with point sources on extended emission
  - in MEM a priori information can be introduced easily through prior image

ME-approach, single aperture

- here a single aperture means that we have *full coverage* of all spatial frequencies, but the S/N ratio drops at scales below the size of the psf *H*
- the maximized function is identical to the previous case except for the inclusion of the point spread function

$$
\begin{aligned}
S &= \int\int f(I)\,dx\,dy \\
&- \lambda\left(\sum_{u,v}\left\{\frac{|\tilde{H}(u,v)\tilde{I}(u,v)-\tilde{D}(u,v)|}{\tilde{\sigma}(u,v)}\right\}^2-\Omega\right)
\end{aligned}
$$

ME-algorithms

- the MEM image is result of optimization
  - optimize $max\{S-\lambda C\}$
  - unknowns are image values $I_{i,j}$ based on which entropy is defined, e.g.,

$$S=-\sum_{i,j} p_{i,j} \log p_{i,j}, \qquad p_{i,j}=I_{i,j}/\sum_{i,j} I_{i,j}$$

  - one can include condition for flux preservation

$$max\{S-\lambda C-\mu \sum I_{i,j}\}$$

    - μ can be selected to fit given value of total flux
    - more simply, use corresponding value *A* directly as the prior

$$S=-\sum_{i,j} I_{i,j}\left[\log(I_{i,j}/A)-1\right]$$

ME-algorithms

- constrained optimization: maximize S subject to constraints on goodness-of-fit C
  - and possible other constraints on flux etc.
- solution is always iterative
- could be found with general optimization algorithms
- there are a number of specialized algorithms

ME-algorithms

– Gull & Daniel (1978) maximize $Q = S - \lambda C$

$$I_j^{(n+1)} = A \exp\left[-\lambda \frac{\partial C(I^{(n)})}{\partial I_j}\right]$$

- image remains positive on all iterations
- *unstable*, even if successive iterates are smoothed

– normal steepest ascent

$$I_j^{(n+1)} = I_j^{(n)} + x \frac{\partial Q(I^{(n)})}{\partial I_j}$$

- develops negative values, unless *x* is extremely small
- negative values must be reset and even then there will be convergence problems

ME-algorithms

- conjugate gradient algorithms
  - instead of direction ∇Q one uses only part that is conjugate to some previous directions (~attempts to estimate the Hessian based on previous steps)
  - considerably better than steepest ascent although problem of negative values persists
- search of unconstrained optimization (λ=const)
  - main expense is in the image data transformations needed for estimates of ∇Q
  - instead of a single line, it may be more efficient to search full subspace
  - some (<10) vectors are used to span the subspace

ME-algorithms

- search directions of constrained problem
  - in previous case a separate iteration is needed on λ so that C becomes correct
  - the search directions e can be selected to enforce the constraints directly
    - e.g. see Skilling & Bryan (1984)

ME-algorithms

- Cornwell & Evans (1985; AIPS task 'VM')
  - maximize $J = S - \alpha\chi^2 - \beta I$
    - Lagrange multipliers α and β selected so that χ^2 and total flux *F* both get their expected values
  - condition $\nabla J=0$ leads to the implicit formula

$$I_i = m_i \exp\left(-\alpha(\partial\chi^2/\partial I_i) - \beta\right)$$

    - *m* is the prior image
  - iterative substitution leads to unstable algorithm and slow convergence (see above) => better to optimize directly the *J*

ME-algorithms

- quadratic approximation of entropy is valid but only very close to solution => need to use second order methods
- direct Newton-Raphson gives

$$\begin{aligned} \Delta I &= (-\nabla\nabla J)^{-1}\nabla J \\ \nabla J &= \nabla S - \alpha\nabla\chi^2 - \beta I \\ \nabla\nabla J &= \nabla\nabla S - 2\alpha H \end{aligned}$$

- Hessian of *J* is diagonal – apart from the contribution from beam profile *H*
- approximation: neglect non-diagonal part of the Hessian = sidelobes of the beam

ME-algorithms

- beam *H* is replaced with a scaled identity matrix

$$\nabla\nabla J = \nabla\nabla S - 2\alpha q I$$

  - *q* is a scaling factor that depends on the beam solid angle (conversion from flux in the beam to flux within a pixel) – exact value is not important
- previous equations give the search direction Δb along which a line search is performed
  - two convolutions are required for the calculation of residuals but also these can be interpolated

ME-algorithms

- convergence criterion $\|\nabla J \cdot \nabla J\| < \epsilon \|1 \cdot 1\|$
- update of lagrange multipliers

$$\begin{aligned} \Delta\alpha &= -\Delta\chi^2 / \|\nabla\chi^2 \cdot \nabla\chi^2\| \\ \Delta\beta &= -\Delta F / \|\nabla F \cdot \nabla F\| \end{aligned}$$

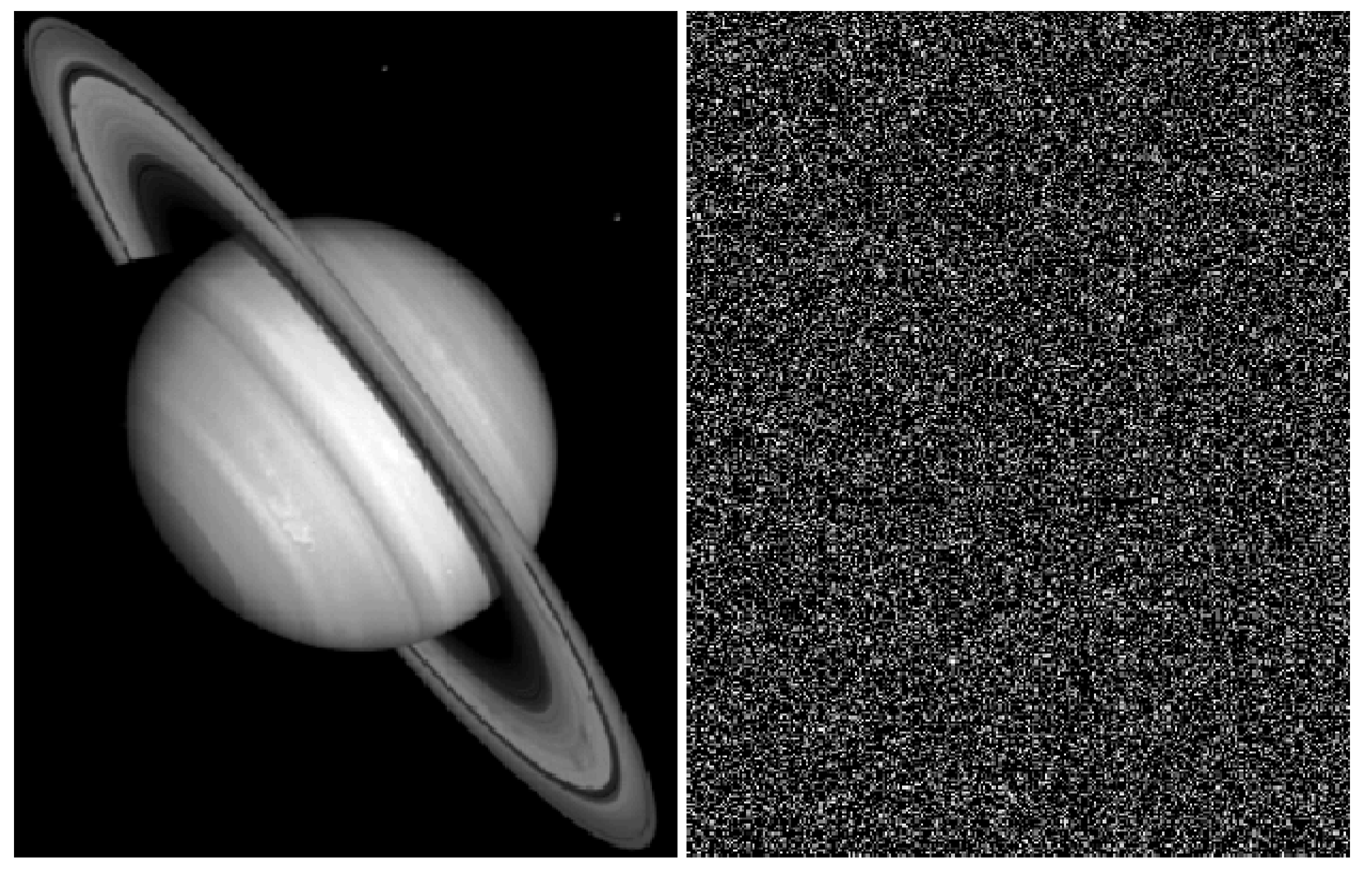
  - these updates interfere with update of *J* so that step size must be limited
- one must correct for negative values
  - small values are cut, lower limit decreased during the iterations
- for large images VM is generally faster than CLEAN

Problems in deconvolution

- Starck & Pantin (2002)
  - fourier based methods give band-limited solutions (Wiener filtering, Tikhonov method, ...)
  - CLEAN cannot restore extended emission (and is slow for large images)
  - MEM cannot recover both compact and entended sources
    - results depend on the background level
    - poor results for features below the background level
    - spatial correlations ignored
  - iterative methods cause noise amplification
    - van Cittert, Richardson-Lucy, Landweber, ...

Problems in deconvolution

– two images with identical entropy



Starck et al. 2001

Wavelet transform

- Fourier-based methods perform poorly when signal contains point sources or edges
  - base functions extend over the whole space
- wavelet transform promises to be a good alternative
  - base functions are wavelets that are localized in both real and frequency space
  - data is presented as a sum of wavelets that are scaled and translated
  - presentation is hierarchical, each level describing the structures at one particular scale

Wavelet transform

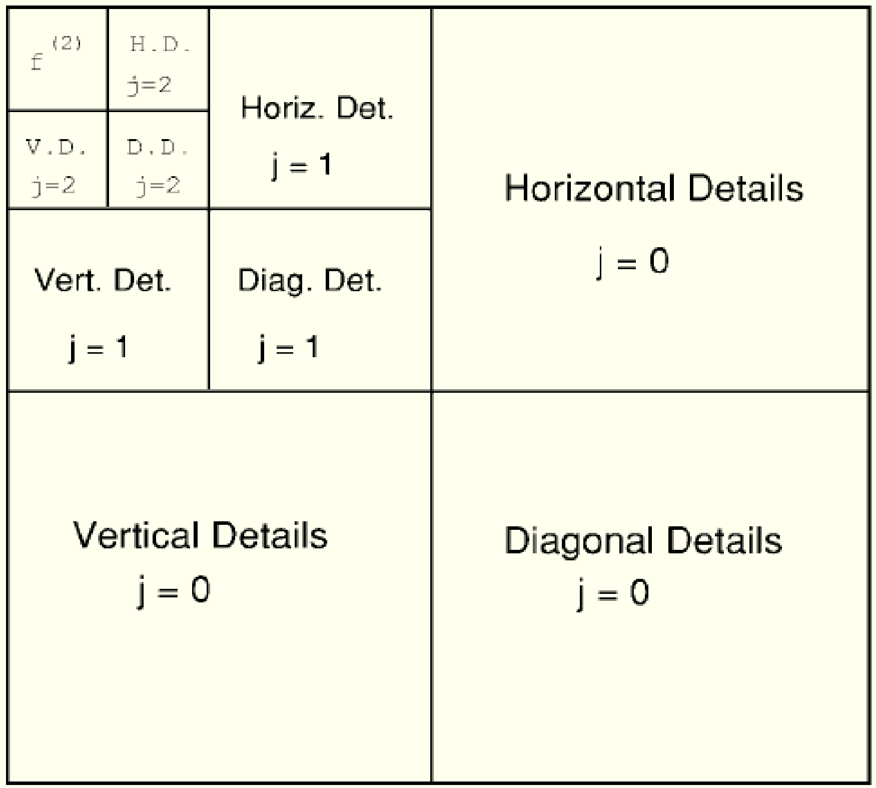
- original signal *s* is decomposed in to a coarse image c_J and wavelet bands w_j , $j=1,\ldots,J$
  - J is the number of scales used
- the coarse image corresponds to frequencies $<(1/2)^J$ and each wavelet band to frequencies $[\,(1/2)^{J+1}, (1/2)^J\,]$
- the decomposition is calculated using low- and high pass filters *h* and *g*

$$c_{j+1,l} = \sum_k h(k-2l)\, c_{j,k}$$
$$w_{j+1,l} = \sum_k g(k-2l)\, c_{j,k}$$

  - filters are derived from the wavelet function

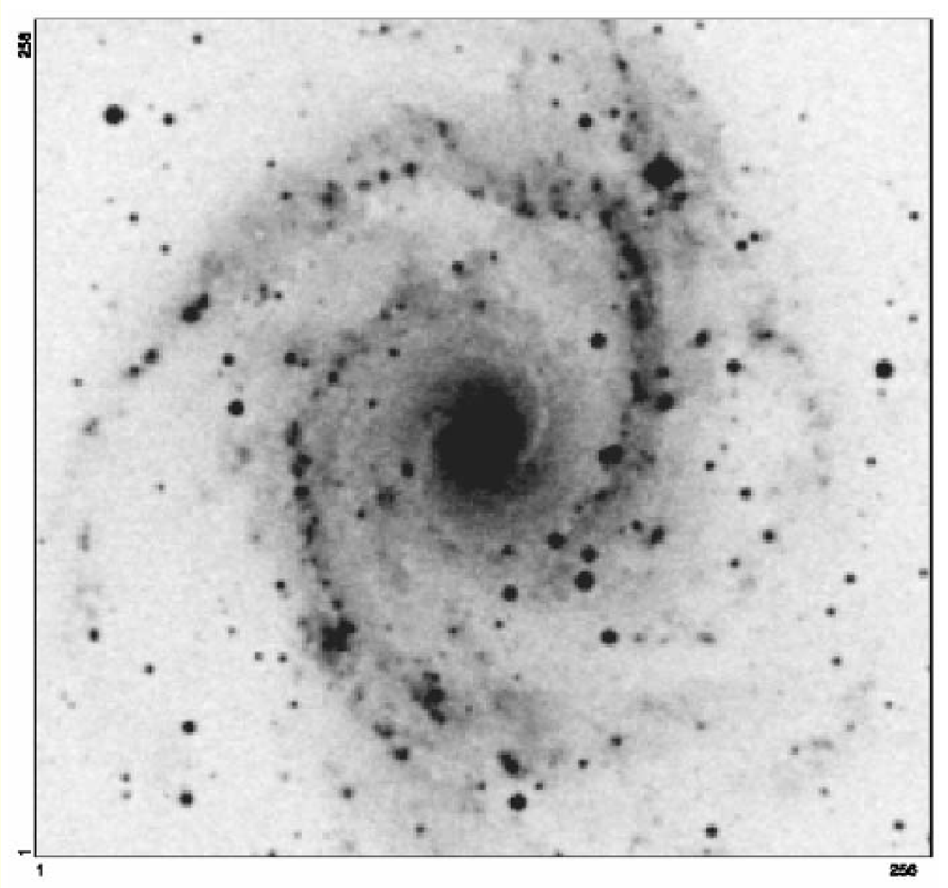
Wavelet transform

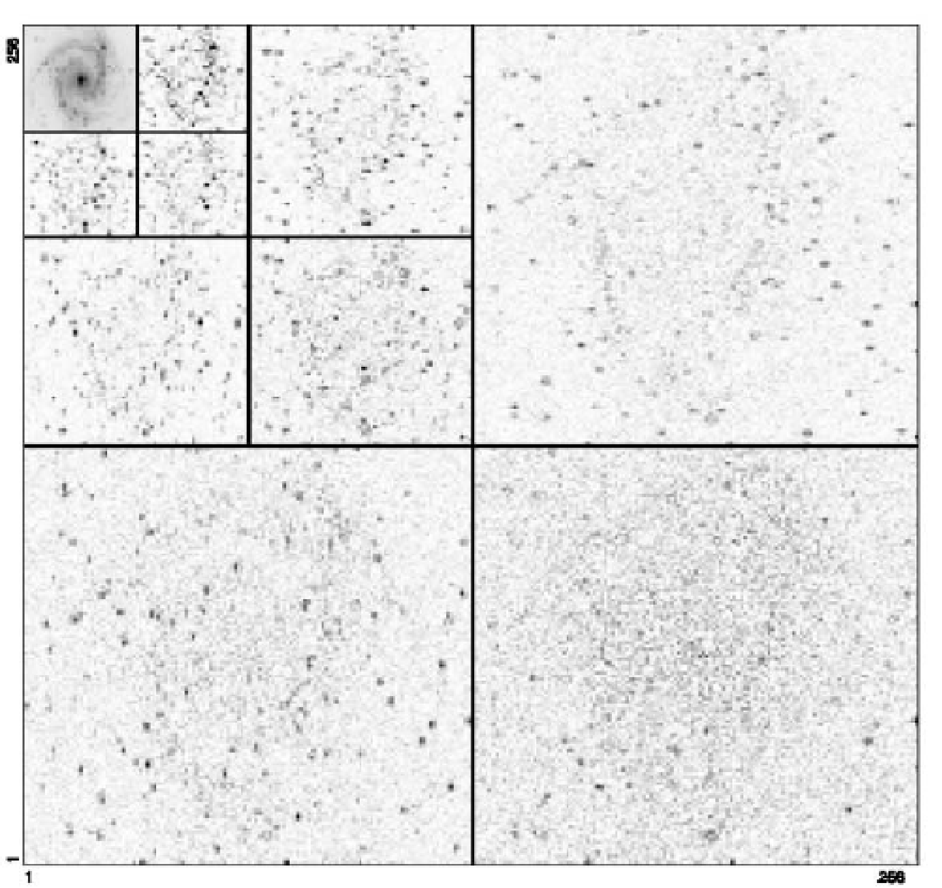
– for 2D images there are three wavelets, one horizontal, one vertical, and one diagonal
  - three wavelet images on each resolution level
  - total number of pixels same as in the original data



Wavelet transform

- NGC 2997





Starck, Pantin, Murtagh (2002)

Wavelet transform

- deconvolution can be done by first applying inverse filter H^{-1}

$$\begin{aligned} \tilde{H}^{-1}(u,v) &= 1/\tilde{H}(u,v) \\ F &= H^{-1} * D + H^{-1} * N = I + Z \end{aligned}$$

  - in *F* the noise *Z* is still normally distributed, and may be **amplified**
  - the wavelet transformation of *F* is **thresholded** and inverse transformation provides the result
    - thresholding sets small wavelet coefficients (=noise) to zero
  - **wavelet-vaguelette** method (Donoho -95)

Wavelet transform

- Neelamani (1999, 2001) **hybrid scheme**
  - regularization still done in Fourier domain through a window function *W*

$$\tilde{W} = \frac{|\tilde{H}|^2}{|\tilde{H}|^2 + \lambda \sigma^2 / \tilde{S}}$$

    - *S* is the noise power spectrum !
  - the windowed function *F* is

$$F = W * H^{-1} * D + W * H^{-1} * N$$

    - parameter λ should be small
    - remaining noise eliminated with wavelet transform (eliminates Gibbs oscillations)
    - positivity constraint not used

Wavelet transform

- some problems of the previous approach
  - determination of the regularization parameter λ is not trivial
  - positivity constraint is not used at all
  - the power spectrum of noise is usually unknown
  - restricted to the case of Gaussian noise

Wavelet regularization

– in iterative deconvolution the residual at a particular iteration is

$$R^{(n)}(x, y) = D(x, y) - (H * I^{(n)})(x, y)$$

– using a wavelet algorithm the residuals can be written as sum of last smooth array and wavelets at J scales

$$R^{(n)}(x, y) = c_J(x, y) + \sum_{j=1}^{J} w_{j, x, y}$$

... but a large part of $w_{j, x, y}$ may be just noise

Wavelet regularization

– need to separate *significant* structures from noise

– define **multiresolution support** *M* as

$$M(j,x,y)=\begin{cases} 1, & if \quad w_j(x,y) \quad significant \\ 0, & if \quad w_j(x,y) \quad insignificant \end{cases}$$

- coefficient is significant if $P(\,|w>w_{j,x,y}|<\varepsilon)$
- for Gaussian noise, e.g., $w>3\sigma_j$
- one can include a source mask in *M*

– one can write 'noiseless' residual

$$\bar{R}^{(n)}(x,y)=c_J(x,y)+\sum_{j=1}^{J} M(j,x,y)\,w_{j,x,y}$$

Wavelet regularization

- with the previous definition one can transform a simple iteration (van Cittert)

$$I^{(n+1)}(x,y) = I^{(n)}(x,y) + \alpha R^{(n)}(x,y)$$

into a more stable scheme

$$I^{(n+1)}(x,y) = I^{(n)}(x,y) + \alpha \bar{R}^{(n)}(x,y)$$

  - only statistically significant structures are carried over to the reconstructed image
  - final result is the restored image and residuals that are pure noise *(R=N)*
- **regularization by significant structures**

Wavelet regularization

– the same can be applied similarly to the Richardson-Lucy algorithm

$$I^{(n+1)} = I^{(n)} \left(\frac{D}{H * I^{(n)}} * H^* \right)$$

– setting $D^{(n)}$ as noiseless data

$$D(x,y) = D^{(n)}(x,y) + R^{(n)}(x,y) \qquad D^{(n)}(x,y) = (H * I^{(n)})(x,y)$$

the regularized version becomes

$$I^{(n+1)} = I^{(n)} \left(\frac{D^{(n)} + \bar{R}^{(n)}}{D^{(n)}} * H^* \right)$$

Wavelet regularization

- the basic idea of the **Pixon** method is somewhat similar to the 'regularization using significant structures'
    - ***not*** based on wavelets; see Dixon et al. 1996
  - – data modelled as a sum of pseudo-images that are smoothed with spatially varying scale
  - – final image consists of a dictionary of features

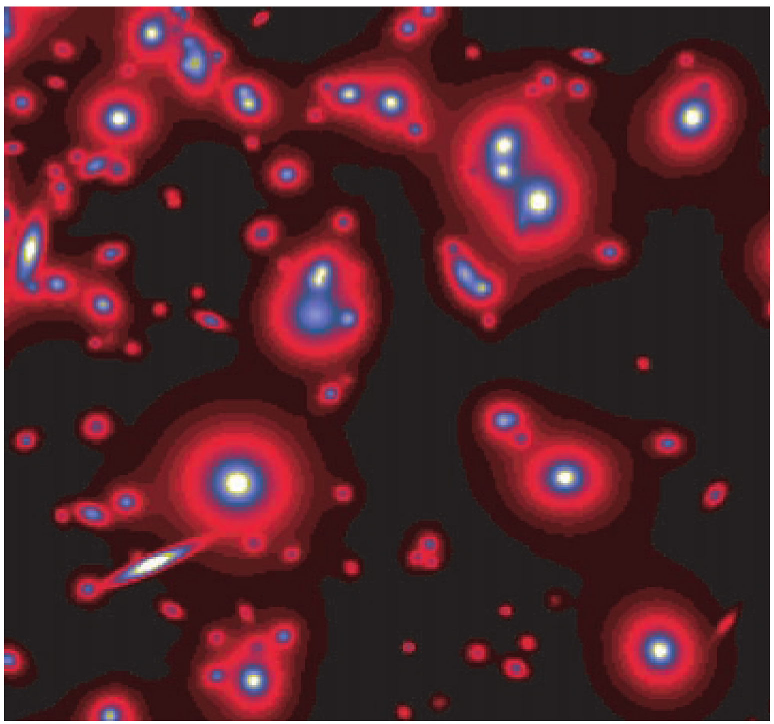
    +weak regularization for strong features

    - if feature cannot be detected directly from data, it is strongly regularized as part of the background

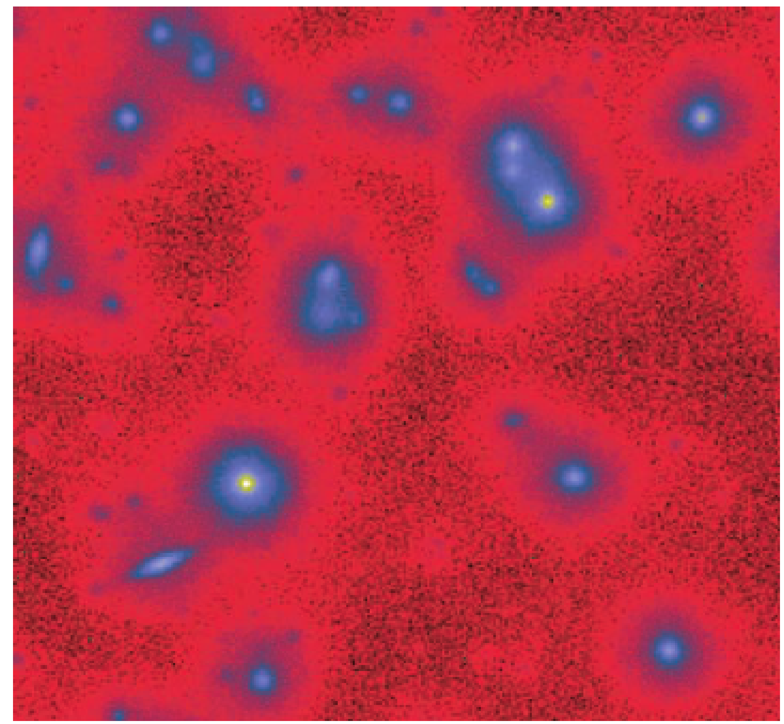
Wavelet regularization

- simulation of galaxy cluster observed with Hubble Wide Field Camera

original: no noise, no aberration



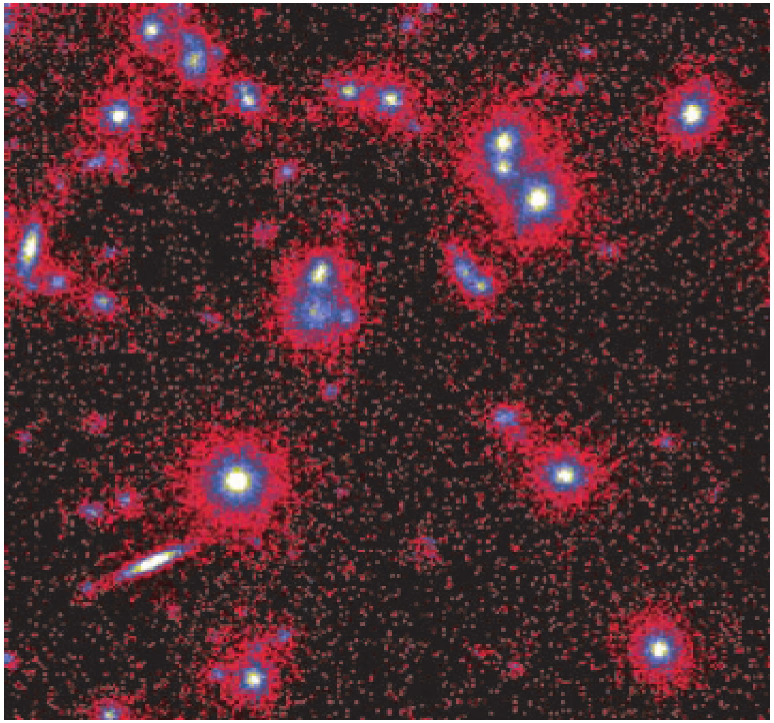
'observed': aberrated and noisy



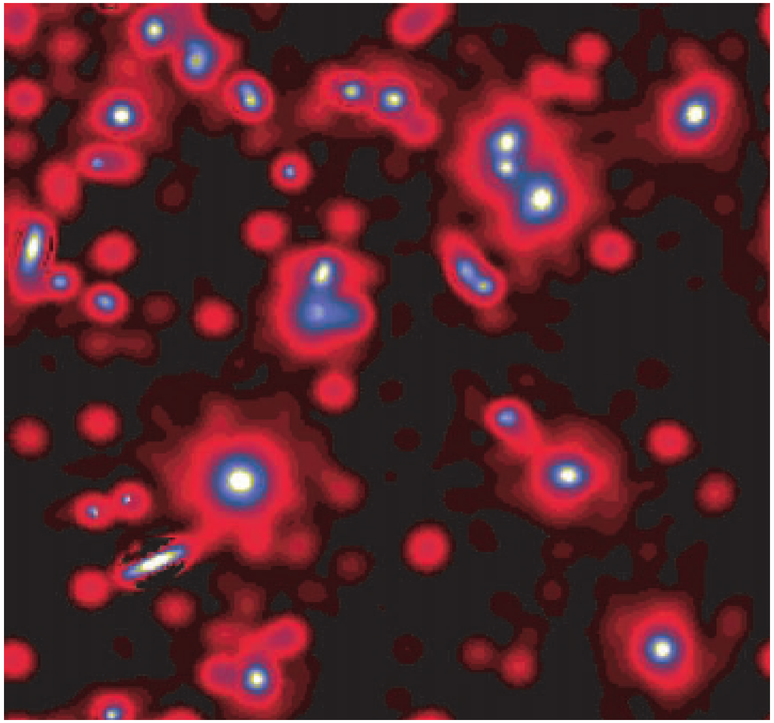
Starck, Pantin, Murtagh (2002)

Wavelet regularization

Richardson-Lucy

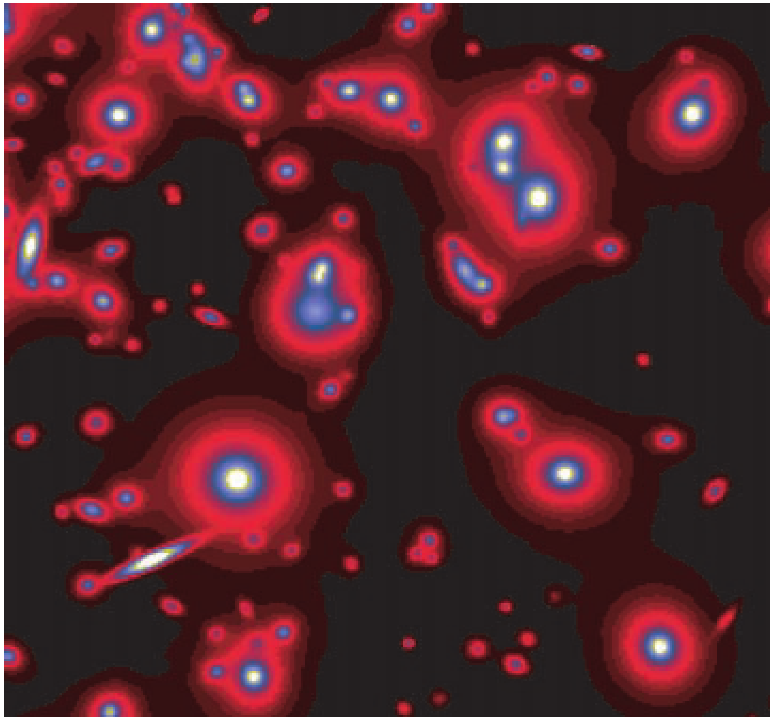


Pixon method

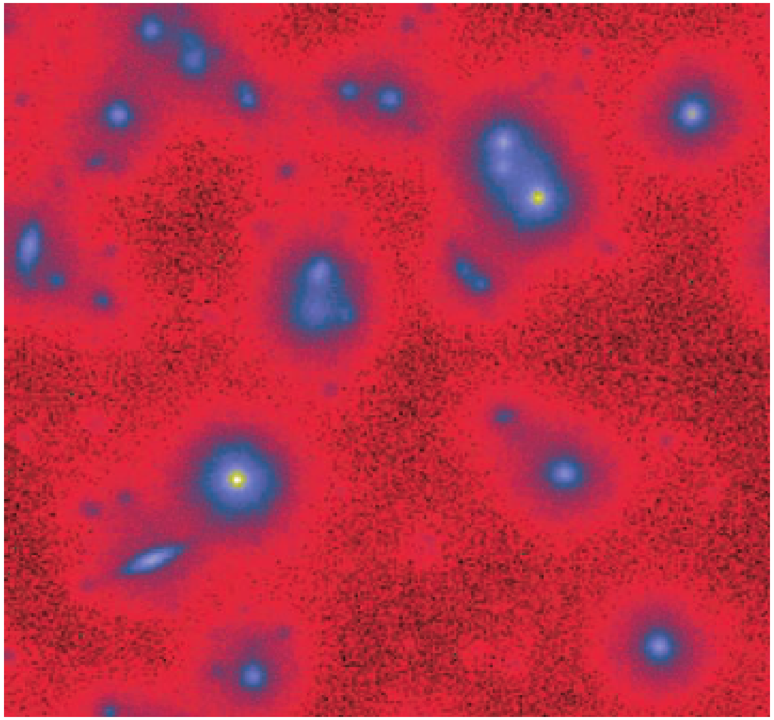


Wavelet regularization

original: no noise, no aberration

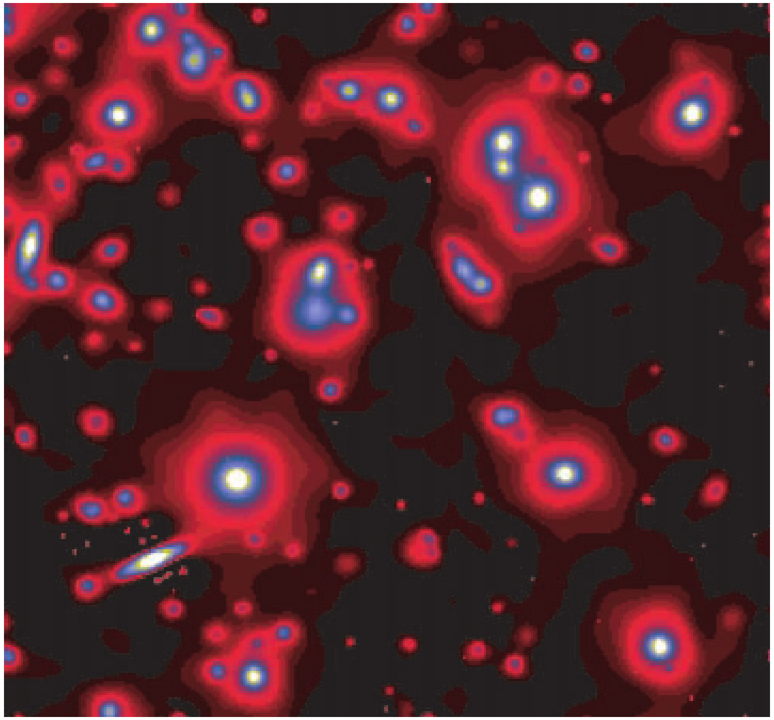


'observed': aberrated and noisy

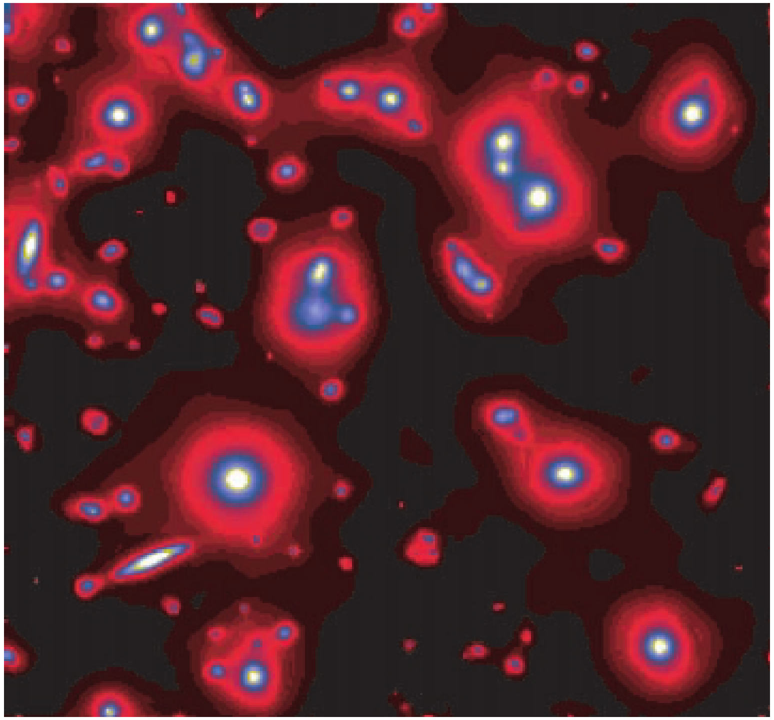


Wavelet regularization

wavelette-vaguelette



wavelet-Lucy



Wavelet regularization

– conclusions from the previous test

- Richardson-Lucy amplifies the noise; some of the fainter objects disappear
- in Pixon method the pixon features are identified from noisy (partially deconvolved data): strong sources are weakly regularized (=good), weak sources suppressed (=bad)
- wavelet-vaguelette is very fast; results better than with the previous methods
- wavelet-Lucy produced best results (fewer spurious sources)

Wavelet CLEAN

- CLEANing
  - consisted of repeated subtraction of strongest *point* sources
  - results were not good for diffuse emission
- MRC CLEAN (Walker & Schwartz 1988)
  - build a smoothed image and the difference between original and the smoothed image
  - apply CLEAN separately to both images
  - result is the sum of processed maps

Wavelet CLEAN

- in wavelet CLEAN the procedure is generalized to many scales (Starck 1998)
  - calculate wavelet transformations of the image, the PSF and the CLEAN beam
  - on each level, apply CLEAN using scale *j* of image and PSF transformations
  - construct result image using the wavelet transformation of the CLEAN beam

Multiscale entropy

- image *I(x,y)* is decomposed into a smooth image $c_{np}(x,y)$ and a set of wavelet coefficients $w_j(x,y)$
  - index *j* refers to the scale, $j=1, \ldots, n_p$
  - original image is $I(x,y)=c_{np}(x,y)+\sum_{j=1}^{np} w_j(x,y)$
  - w_j are calculated as the difference between the last smoothed plane and an image obtained with a low pass filter

$$c_j(k) = \sum_l h(l) c_{j-1}(k+2^{j-1} l)$$

$$w_j(k) = c_{j-1}(k) - c_j(k)$$

Multiscale entropy

– entropy is related to the sum of information on each scale

– in wavelet transformation information is related to the probability that individual wavelet coefficients are caused by noise

$$S = -\sum_{j=1}^{J} \sum_{k=1}^{N_j} \ln p(w_{j,k})$$

– minimize

$$J = \frac{1}{2} \sum_{k=1}^{N} \left[\frac{D_k - (H * I)_k}{\sigma} \right]^2 \pm \alpha S$$

Multiscale entropy

– for Gaussian noise

$$S = \sum_{j=1}^{J} \sum_{k=1}^{N_j} s(w_{j,k})$$

where

$$s(w_{j,k}) = \frac{w_{j,k}^2}{2\sigma_j^2} + constant$$

- s_j is again the noise in the wavelet coefficients at the level j
- the constant term does not affect the maximization
- entropy ~ information => with the above definition smooth solution requires **minimization** of entropy

Multiscale entropy

- in Starck & Pantin (1996) multiscale entropy is defined

$$S_m = \frac{1}{\sigma_I} \sum_{scales\ j} \sum_{pixels} \sigma_j \left[w_j(x,y) - m_j - |w_j(x,y)| \ln \frac{|w_j(x,y)|}{m_j} \right]$$

  - in the absence of signal the wavelet coefficients at level *j* approach m_j – this should be small compared with any real signal, i.e., a small fraction of the noise
  - the noise σ_j acts as a weighting factor of the different scales
  - entropy **maximized**

Multiscale entropy

- multiresolution support
  - images at all wavelets scales, with pixel value TRUE if some information is present at given scale and location (hard threshold, 'hard weighting')

1. calculate wavelet transform of image *I(x,y)*
2. estimate σ of the scale and set a threshold value for the significance (e.g. 3σ)
3. multiresolution support *M(j,x,y)* is

$$M(j,x,y)=\begin{cases} 1, & if \quad w_j(x,y)\geqslant k\sigma_j \\ 0, & if \quad w_j(x,y)<k\sigma_j \end{cases}$$

Multiscale entropy

- the aim
  - to reconstruct significant structures (where we have enough signal) without strong regularization
  - to eliminate noise (strong regularization)
- new definition for multiscale entropy

$$S_m = \frac{1}{\sigma_I} \sum_{scales\ j} \sum_{pixels} \left[1 - M(j,x,y)\right] \sigma_j \left[w_j(x,y) - m_j - |w_j(x,y)| \ln \frac{|w_j(x,y)|}{m_j} \right]$$

  - entropy is calculated only for scales and regions with low signal-to-noise ratio

Multiscale entropy

- previous M can be replaced with a continuous function going from 0 to 1
  - $M \sim 1$ => weak regularization
  - $M \sim 0$ => strong regularization
- the values σ_j can be obtained from simulations
  - create noise image with $\sigma=1$ and do wavelet transform
  - calculate standard deviations of wavelet coefficients at each scale, σ_j^e
  - $\sigma_j = \sigma_I \sigma_j^e$

Multiscale entropy

- image noise σ_I
  - estimated from the difference of the image and average filtered image
  - e.g. in the case of CCD images one should disregard the high noise borders
    - look at the intensity histogram
  - estimate can be improved iteratively, using the wavelet transformation
    - calculate multiresolution support and re-estimate noise from pixels $M \sim 0$

Multiscale entropy

- for minimization of

$$J=\sum_{pixels}\frac{1}{2}\left(\frac{D-H*I}{\sigma_I}\right)^2-\alpha S_{ms}(I)$$

Starck & Pantin (1996) used the gradient

$$\nabla J=-H^{*}*\ \frac{D-H*I}{\sigma_I^2}+\frac{\alpha}{\sigma_I}\sum_{scale\ j}\left[\left[1-M(j)\right]\sigma_j\, sgn(w_j^{(0)})\ \ln\left(\frac{|w_j^{(0)}|}{m_j}\right)\right]*\Psi_j^{*}$$

and steepest descent $I^{n+1}=I^n-\gamma\nabla J(I^n)$

- above Ψ_j is the wavelet at scale *j*

Multiscale entropy

- the parameter α is fixed by the noise behaviour of the wavelet transform
  - determination far from straightforward
  - depends on the requirement of smoothness

Multiscale entropy

- Starck et al. (2001)
  - the coefficient *w* is partly due to signal (S_S), partly due to noise (S_N)
    - small coefficients are noise and contribute to S_N
    - large coefficients contribute mostly to S_S
  - one might want to minize

$$J = \frac{1}{2} \sum_{k=1}^{N} \left[\frac{D_k - (P * I)_k}{\sigma} \right]^2 + \alpha S_N(I)$$

    - in the solution minimize the information that is *due to the noise*

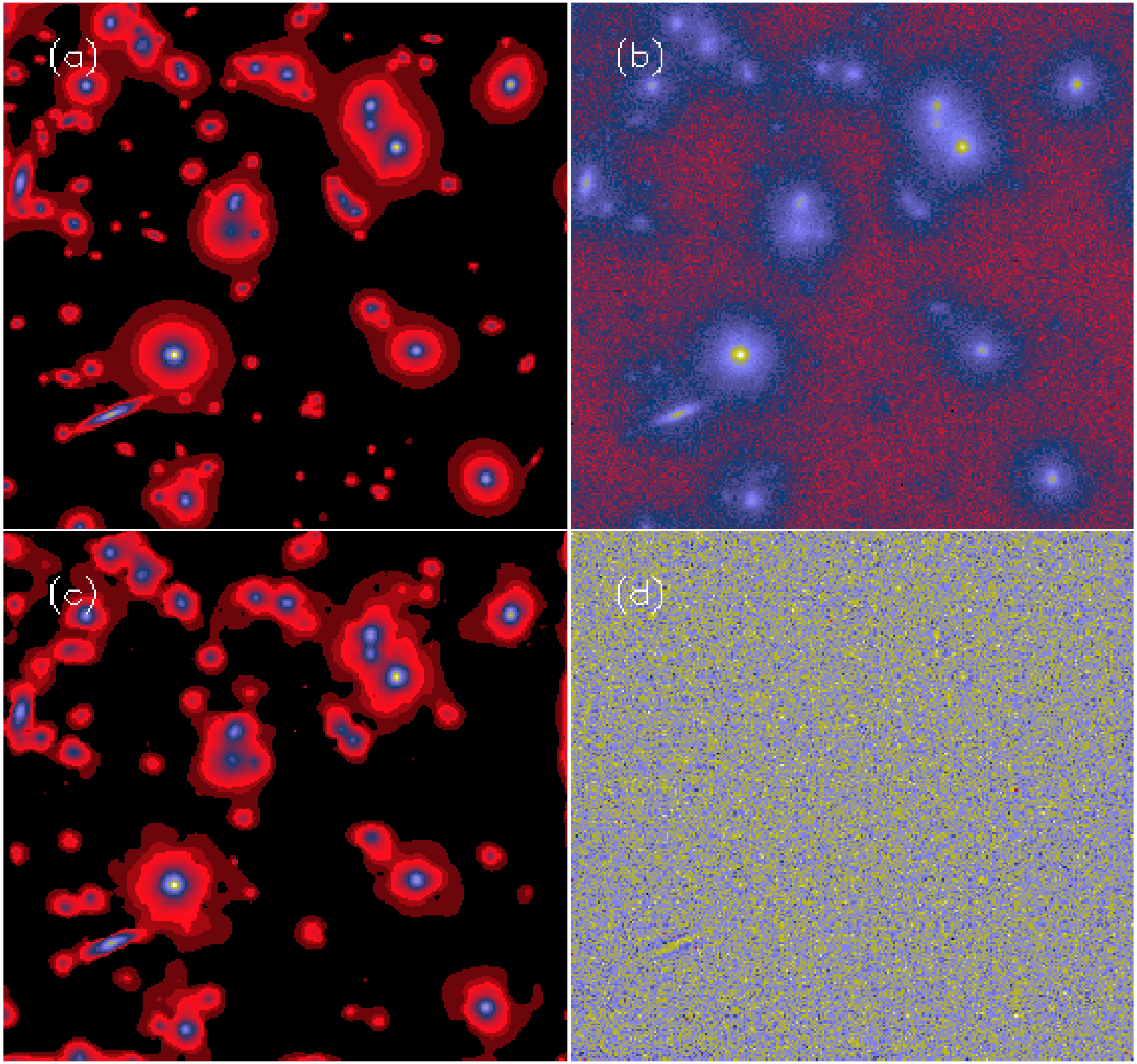
Multiscale entropy

- the minimized function can be generalized as

$$J = S_S(D - P * I) + \alpha S_N(I)$$

  - minimize signal contribution to the information in the residuals
  - minimize the noise contribution to the information in the solution
- both hard and soft weighting were studied, with and without the regularizing term

Multiscale entropy



a) original
b) blurred and with Gaussian noise
c) de-convolved image
d) residuals

Starck et al. (2001)

Multiscale entropy

– multiscale entropy can be used to test presence of undetected sources

- with $h(w_{j,k}) = -\ln[p(w_{j,k})]$, calculate mean entropy of each scale j

$$E(j)=\frac{1}{N}\sum_{k=1}^{N_j} s(w_{j,k})$$

- for assumed noise model, calculate normalized mean entropy

$$E_n(j)=\frac{E(j)}{E^{(noise)}(j)}$$

– $E^{(noise)}$ can be calculated ... at least with simulations

Multiscale entropy

- plot normalized entropy as the function of scale
- weak sources increse entropy at small scales, even when individual objects cannot be detected

– simulation (Starck 2001):

- 0, 50, 100, 200, or 400 sources with maximum equal to noise σ and source standard deviation 2

Multiscale entropy

J.-L. Starck et al.: Entropy and astronomical data analysis 743

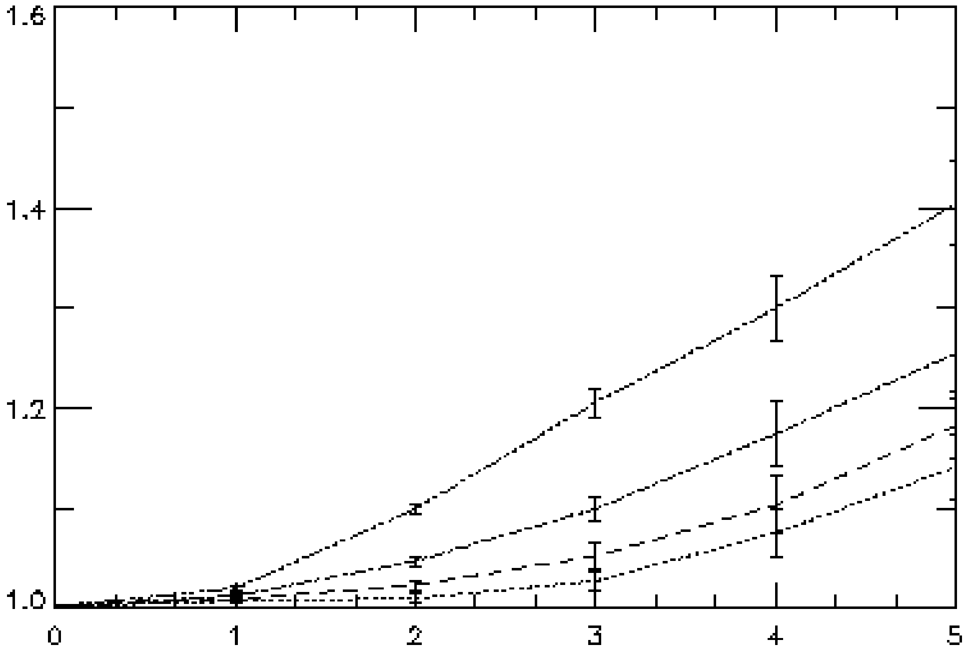


Fig. 9. Mean entropy versus the scale of 5 simulated images containing undetectable sources and noise. Each curve corresponds to the multiscale transform of one image. From top to bottom, the image contains respectively 400, 200, 100, 50 and 0 sources

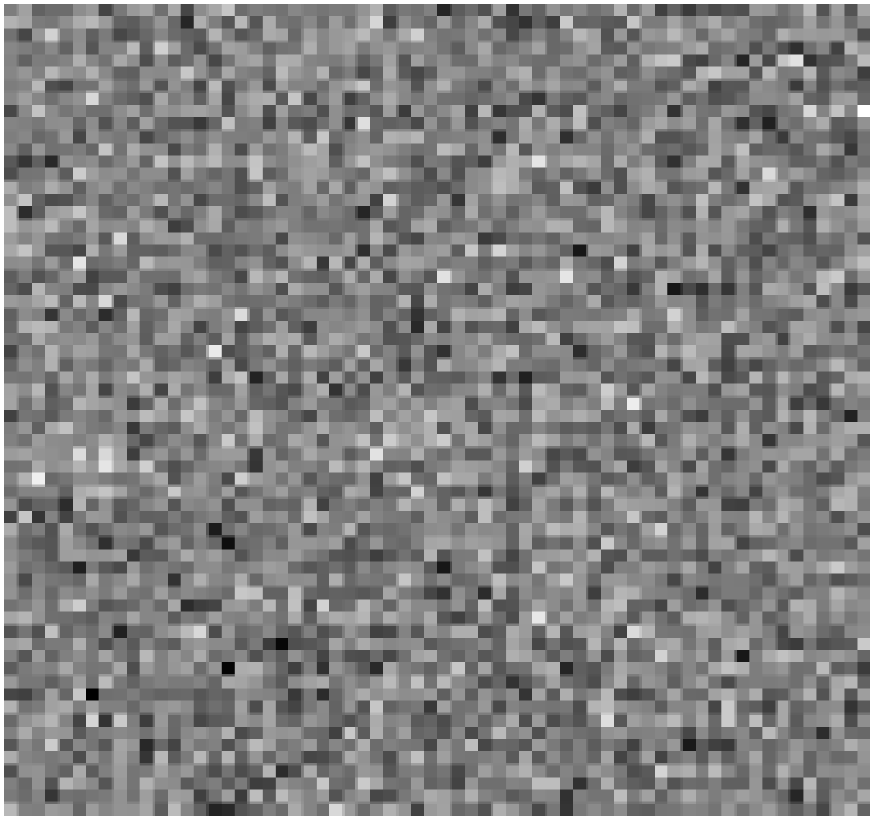


Fig. 10. Region of a simulated image containing an undetectable source at the center

Superresolution

- superresolution means resolution better than the beam size (diffraction limit)
- two conditions
  - space-bandwidth product (SBP) is no more than ~1
    - SBP is approximately the ratio between source diameter and the diffraction limit
    - => source must be small
  - good signal to noise ratio (SNR)

Superresolution

- method requires *some* information about the source model
  - flux of the reconstructed object is concentrated within a smaller area
  - restoration is stable, if the total flux of the object is constrained
  - even non-negativity of the MEM can produce superresolution (of 'nearly black objects')
  - Richardson-Lucy implements both non-negativity and flux conservation

Superresolution

- data is kept at original resolution
- model and the PSF are sampled at higher resolution
- for example, in the case of MAP Poisson algorithm

$$I^{(n+1)} = I^{(n)} \exp\left\{ \left(\frac{D}{\left(H * I^{(n)}\right)_{\downarrow}} - 1 \right)_{\uparrow} * H^{*} \right\}$$

  - oversampling = '↑', down-sampling = '↓'

Superresolution

- PSF undersampled but observations are made with small shifts => deconvolved image can be reconstructed on finer grid
  - co-adding of frames on fine grid: operator $L_{\uparrow}$
  - estimation of k^{th} observation from $L_{\uparrow}D$: operator $L^{-1}_{\downarrow}$
  - Landweber iteration becomes

$$I^{(n+1)} = I^{(n)} + \alpha H^{*} \left[L_{\uparrow} \left(D - L_{\downarrow}^{-1} \left(H * I^{(n)} \right) \right) \right]$$

    - PSF function H is needed at the fine resolution

Superresolution

- example: Anconelli (2005)
  1. apply Richardson-Lucy
  2. if stars are resolved, go to step 3. Otherwise
     1. define domain D where source intensity above threshold
     2. apply RL to that domain
     3. if stars are separated, positions are obtained as the feature centroids
  3. fit a least squares model (sum of psf's)
     - ... in case feature still contains several stars

Superresolution

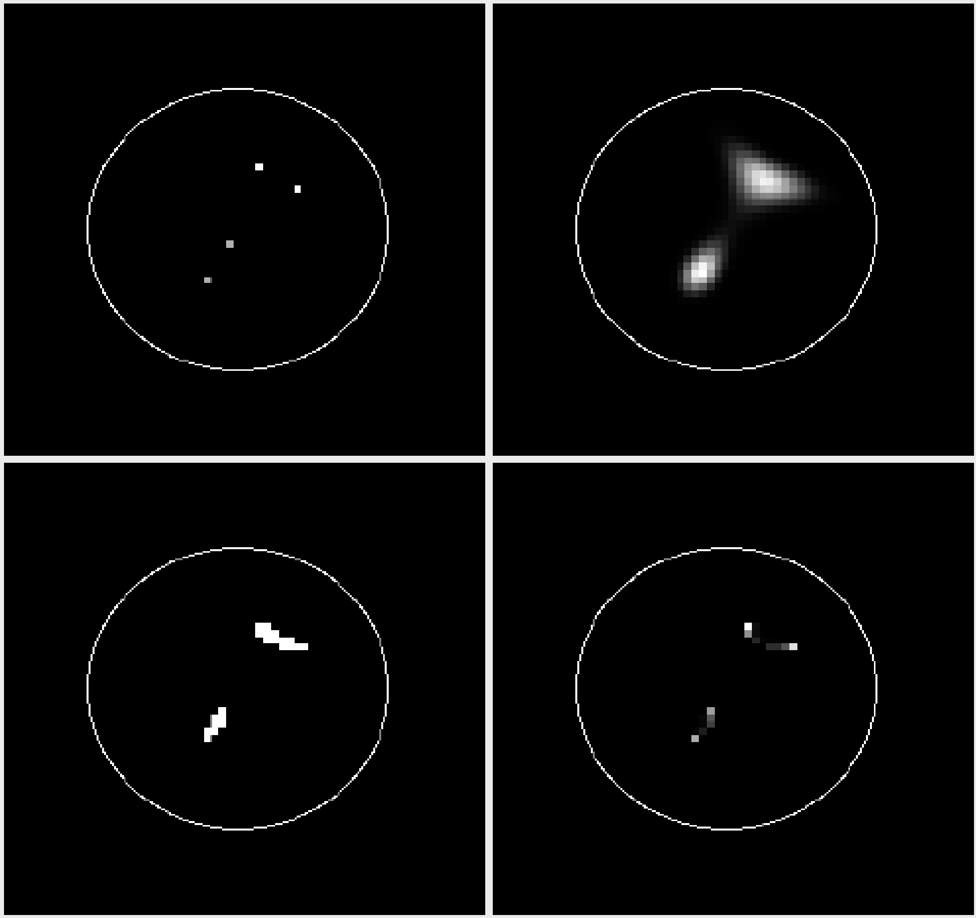
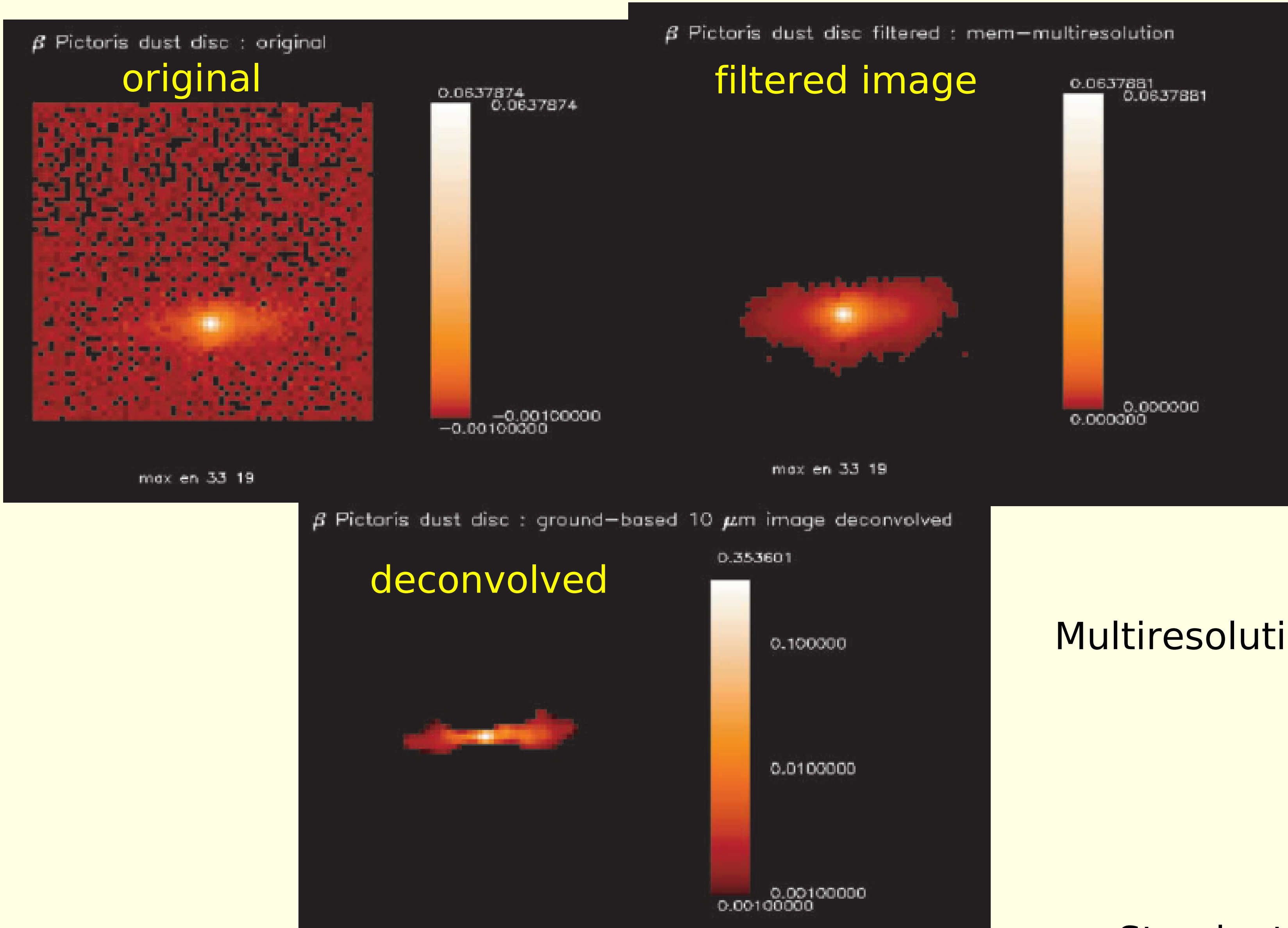


Fig. 1. *Upper-left panel*: the object; *upper-right*: the reconstruction after 1000 OS-EM iterations (first step); *lower-left*: the mask obtained with a thresholding of the previous reconstruction (50% of the maximum value); *lower-right*: the restoration after 1000 OS-EM iterations, initialized with the previous mask (second step). The diameter of the circle shown in each picture is λ/B.

Superresolution



Multiresolution MEM

Starck et al. 2002

Program packages

- IRAF/STSDAS
  - Lucy (PLucy?); `lucy`
  - maximum entropy: `mem`

- Starlink, package *Kappa*
  - Wiener filtering
  - Richardson-Lucy
  - maximum entropy

Program packages

- MIDAS
  - core commands
    - DECONVOLVE/IMAGE frame psf result [no_iter] [cont_flag]
      -
  - package surfphot
    - REBIN/DECONVOLVE frame psf result, zoom_x, zoom_y, n_iter
      -

Program packages

- Midas ctd.
  - package wavelet
    - FILTER/WAVE
      - thresholding, multiresolution Wiener filtering
    - GRAD/WAVE
      - deconvolution by regularized one step algorithm
    - LUCY/WAVE
      - wavelet Lucy
    - DIRECT/WAVE
      - multiresolution Tikhonov, regularization term $\gamma_j \|w_j\|^2$
    - CITTERT/WAVE
      - van Cittert + multiresolution support

Program packages

- MR/1-2 by Starck & Murtagh
  - commercial program (www.multiresolution.com)
  - free version restricted to max 256x256 images
  - standard deconvolution methods (MEM, Lucy, Landweber, MAP, ...)
  - wavelet based methods
    - including multiresolution MEM

References

- Anconelli et al., 2005, Restoration of interferometric images: III. Efficient Richardson-Lucy methods..., A&A 430, 731
- Anconelli et al. 2005, Restoration of interferometric images: IV. An algorithm for super-resolution of stellar systems, A&A 431, 747
- Charbonnier et al., 1997, IEEE Trans. Image Process. 44, 2988
- Cornwell, Evans: A simple maximum entropy deconvolution algorithm, 1985, A&A 143, 77
- Dixon et al. 1996, A&AS 120, 683

References

- Donoho: Nonlinear solution of inverse problems by wavelet-vaguelette decomposition, 1995, Applied and Computational Harmonic Analysis 2, 101
- Lucy: An iterative technique for the rectification of observed distributions, 1974, AJ 79, 745
- Moline et al. 1992, AJ 104, 1662
- Lucy: Optimum strategies for inverse problems in statistical astronomy, 1994, A&A 289, 983
- Midas Users Guide, `www.eso.org/esomidas`
- Molina et al. 2001, IEEE Signal Process. Mag. 18, 11
- Narayan, Nityananda: Maximum entropy image restoration, 1986, ARA&A 24, 176

References

- Neelamani: Wavelet based deconvolution for ill-conditioned systems, MS thesis, 1999, Rice University
- Nunez, Llacer: A general bayesian image reconstruction algorithm with entropy prior, 1993, PASP 105, 1192
- Pina, Puetter: Incorporation of spatial information in Bayesian image reconstruction: the maximum residual likelihood criterion, 1992, PASP 104, 1096
- Pina, Puetter: Bayesian image reconstruction: the pixon and optimal image modelling, 1993, PASP 105, 630
- Roumeliotis: A novel maximum likelihood method for image reconstruction with increased sampling, 1995, ApJ 452, 944

References

- Skilling, Bryan: Maximum entropy image reconstruction: general algorithm, 1984, MNRAS 211, 111
- Starck, Murtagh, Bijaoui, 1998; Image Processing and Data Analysis: The Multiscale Approach, CUP
- Starck et al.: Entropy and astronomical data analysis: Perspectives from multiresolution analysis, 2001, A&A 368, 730
- Starck, Pantin, Murtagh: Deconvolution in astronomy: A Review, 2002, PASP 114, 1051
- Starck, Pantin: Multiscale maximum entropy images restoration, in Vistas in astronomy, Vol. 40 no 4, p. 563

References

- Starck: Multiresolution and its Applications: An Overview, http://thames.cs.rhul.ac.uk/~multires/doc/doc_demo.pdf
- Deconvolution and blind deconvolution in astronomy, http://jstarck.free.fr/Blind07.pdf
- Starck, Murtagh, Bijaoui: Image Processing and Data Analysis: The Multiscale Approach, CUP 1998, http://www.multiresolution.com/cupbook.html
- Terebizh: Superresolution in pattern recognition and image restoration problems, 1993, A&A 270, 543
- Thompson, Craig: Automatic strategies for astrophysical inverse problems, 1992, A&A 262, 359

References

- Tikhonov 1987, Soviet Physics – Doklady 32, 456
- Wu: MEM package for image restoration in IRAF, in ADASS-II, ASP conference series Vol. 52 (1993) p. 520