L12: Model Fitting

EuroSummer School

Observation and data reduction with the Very Large Telescope Interferometer

Goutelas, France
June 4-16, 2006

Eric Thiébaut, Isabelle Tallon-Bosc, Clémentine Béchet, Michel Tallon
Centre de Recherche Astronomique de Lyon / Jean-Marie Mariotti Center
14 June 2006

The problem at hand

what you have:

- data, for instance OI-FITS:
  - OI_VIS complex visibility (amplitude and phase)
  - OI_VIS2 squared visibility amplitude
  - OI_T3 triple product a.k.a. bispectrum (amplitude and phase)
- priors (*i.e.* possible models of the observed object)

what you want:

- identify the observed object
- estimate object parameters and uncertainties

what you need:

- tools for model fitting
- know what you are doing (*no black magic!*)

Other questions

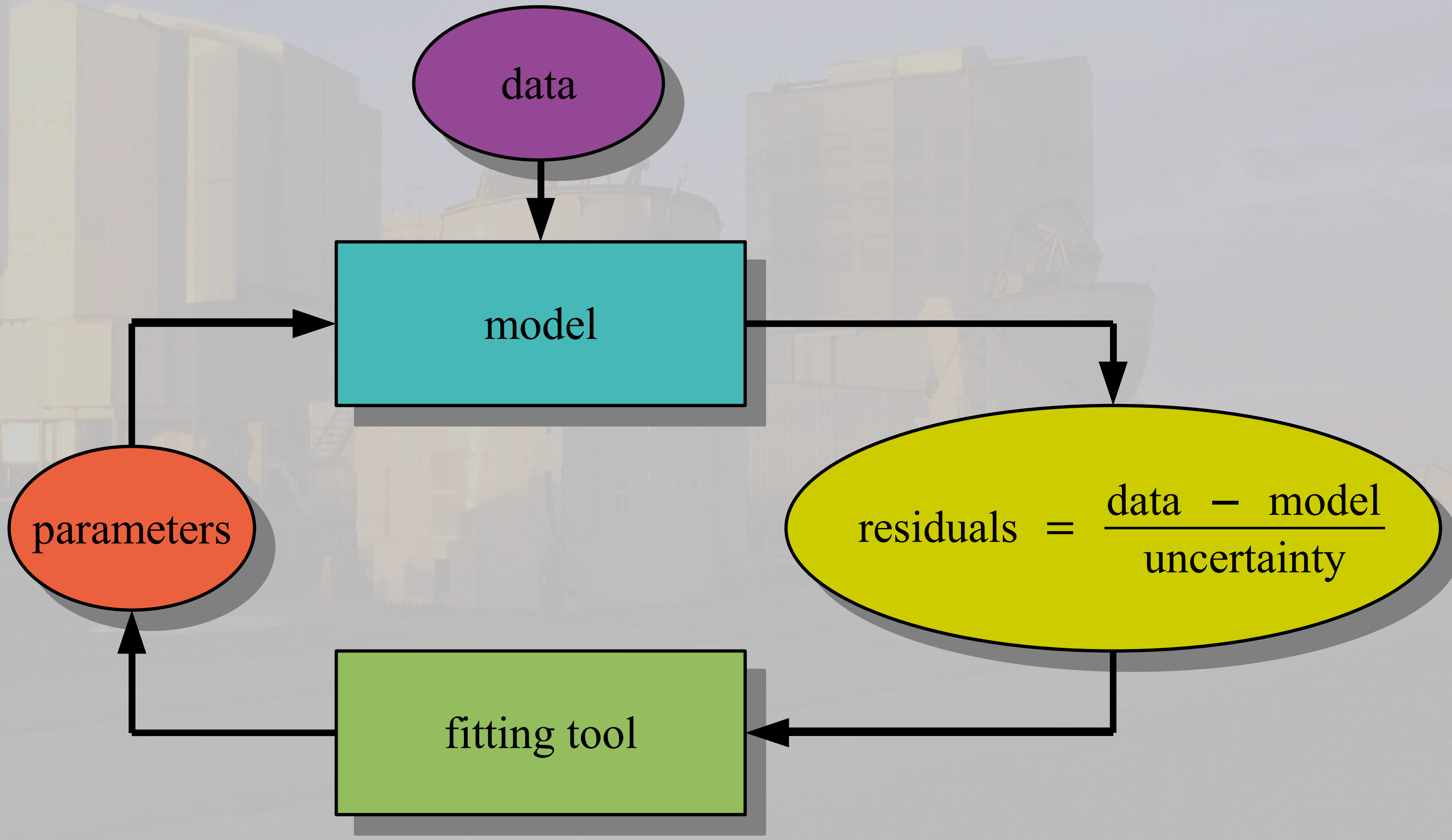
what can you do from optical interferometric data?

- direct interpretation (for gurus!)
- data processing to have human readable view of the data
- image (possibly poly-chromatic) however
  - image may require more measurements (about as many as resels in the synthesized field of view, in fact this is not true but this is another story)
- estimate parameters of a model

what is a *model*?

- a mathematical/numerical function which can predict the data values given the parameters

Model fitting: schematic view



What are the *best* parameters?

- the ones which maximize the probability of having observed the data:

$$\boldsymbol{x}_{\text{best}} = \arg\max_{\boldsymbol{x}} \Pr\left(\boldsymbol{d} \middle| \boldsymbol{m}(\boldsymbol{x})\right)$$

where:

$\boldsymbol{x}$	are the parameters
$\boldsymbol{m}(\boldsymbol{x})$	is the model
$\boldsymbol{d}$	are the data

similarly: $\boldsymbol{x}_{\text{best}} = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$

where $f(\boldsymbol{x}) \propto -\log \Pr\left(\boldsymbol{d} \middle| \boldsymbol{m}(\boldsymbol{x})\right)$

Gaussian Statistics

for Gaussian errors (noise + model errors) and the correct model:

$$\Pr\left(\boldsymbol{d}\middle|\boldsymbol{m}(\boldsymbol{x})\right) = \frac{\exp\left(-\frac{1}{2}\,\boldsymbol{r}^{\mathrm{T}}\cdot\boldsymbol{C}^{-1}\cdot\boldsymbol{r}\right)}{\sqrt{(2\,\pi)^{N_{\mathrm{data}}}\,\det(\boldsymbol{C})}}$$

where the ***residuals*** are:

$$\boldsymbol{r} = \pm\left(\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{x})\right)$$

$\boldsymbol{m}(\boldsymbol{x})$ is the model, $\boldsymbol{d}$ are the data, and $\boldsymbol{C}$ is the ***covariance*** matrix of the residuals:

$$\boldsymbol{C} = \langle \boldsymbol{r}\cdot\boldsymbol{r}^{\mathrm{T}}\rangle - \langle \boldsymbol{r}\rangle\cdot\langle \boldsymbol{r}\rangle^{\mathrm{T}}$$

Gaussian log-likelihood

general Gaussian statistics: $$\Pr(\boldsymbol{r}) = \frac{\exp\left(-\frac{1}{2}\,\boldsymbol{r}^{\mathrm{T}}\cdot\boldsymbol{C}^{-1}\cdot\boldsymbol{r}\right)}{\sqrt{(2\,\pi)^{N_{\mathrm{data}}}\,\det(\boldsymbol{C})}}$$

taking: $$\begin{aligned} f(\boldsymbol{x}) &= -2\,\log\Pr\left(\boldsymbol{d}|\boldsymbol{m}(\boldsymbol{x})\right) \\ &= \left(\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{x})\right)^{\mathrm{T}}\cdot\boldsymbol{C}^{-1}\cdot\left(\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{x})\right) + \cancel{N_{\mathrm{data}}\,\log(2\,\pi) + \log\det(\boldsymbol{C})} \end{aligned}$$

and discarding irrelevant additive constants, yields:

$$\begin{aligned} f(\boldsymbol{x}) = \chi^2(\boldsymbol{x}) &= \left(\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{x})\right)^{\mathrm{T}}\cdot\boldsymbol{C}^{-1}\cdot\left(\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{x})\right) \\ &= \boldsymbol{r}(\boldsymbol{x})^{\mathrm{T}}\cdot\boldsymbol{C}^{-1}\cdot\boldsymbol{r}(\boldsymbol{x}) \end{aligned}$$

$$\boldsymbol{r}(\boldsymbol{x}) = \pm\left(\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{x})\right)$$

$$\boldsymbol{C} = \langle\boldsymbol{r}\cdot\boldsymbol{r}^{\mathrm{T}}\rangle - \langle\boldsymbol{r}\rangle\cdot\langle\boldsymbol{r}\rangle^{\mathrm{T}}$$

Gaussian log-likelihood for independent data

independent Gaussian data: $$\Pr(\boldsymbol{r}) = \prod_{j}^{N_{\text{data}}} \Pr(r_j) = \prod_{j}^{N_{\text{data}}} \frac{\exp\left(\frac{-r_j^2}{2\,\sigma_j^2}\right)}{\sqrt{2\,\pi\,\sigma_j}}$$

$$\chi^2(\boldsymbol{x}) = \sum_{j}^{N_{\text{data}}} \frac{r_j^2(\boldsymbol{x})}{\sigma_j^2} = \sum_{j}^{N_{\text{data}}} e_j^2(\boldsymbol{x})$$

residuals: $$r_j(\boldsymbol{x}) = \pm\left(d_j - \boldsymbol{m}_j(\boldsymbol{x})\right)$$

normalized errors: $$e_j(\boldsymbol{x}) = \frac{r_j(\boldsymbol{x})}{\sigma_j}$$

Statistics of complex data

complex visibility: $z = x + \mathrm{i}\, y = \rho\, \mathrm{e}^{\mathrm{i}\,\phi}$

$$x = \rho \cos\phi$$
$$y = \rho \sin\phi$$

complex residuals: $r = \delta z = \delta x + \mathrm{i}\, \delta y$

chi-square (2 possible expressions!):

$$\chi^2 = \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}^{\mathrm{T}} \cdot \begin{pmatrix} \sigma_x^2 & C_{x,y} \\ C_{x,y} & \sigma_y^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = \frac{1}{\sigma_x^2\, \sigma_y^2 - C_{x,y}^2} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}^{\mathrm{T}} \cdot \begin{pmatrix} \sigma_y^2 & -C_{x,y} \\ -C_{x,y} & \sigma_x^2 \end{pmatrix} \cdot \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}$$

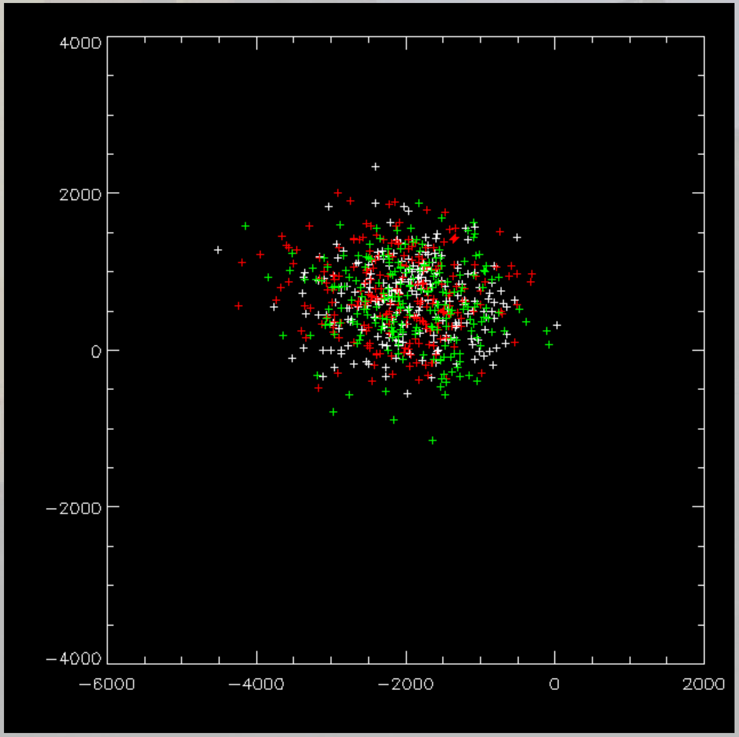
$\neq$

$$\chi^2 = \begin{pmatrix} \delta \rho \\ \delta \phi \end{pmatrix}^{\mathrm{T}} \cdot \begin{pmatrix} \sigma_\rho^2 & C_{\rho,\phi} \\ C_{\rho,\phi} & \sigma_\phi^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \delta \rho \\ \delta \phi \end{pmatrix}$$

← amplitude and phase residuals

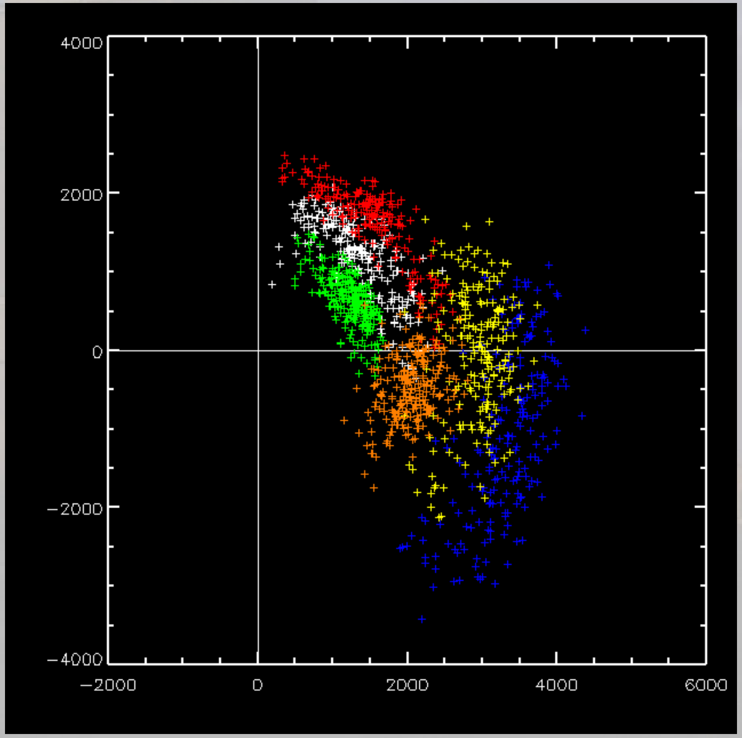
Statistics of *real* optical interferometry data

low SNR



(triple product of FKV0509)

high SNR



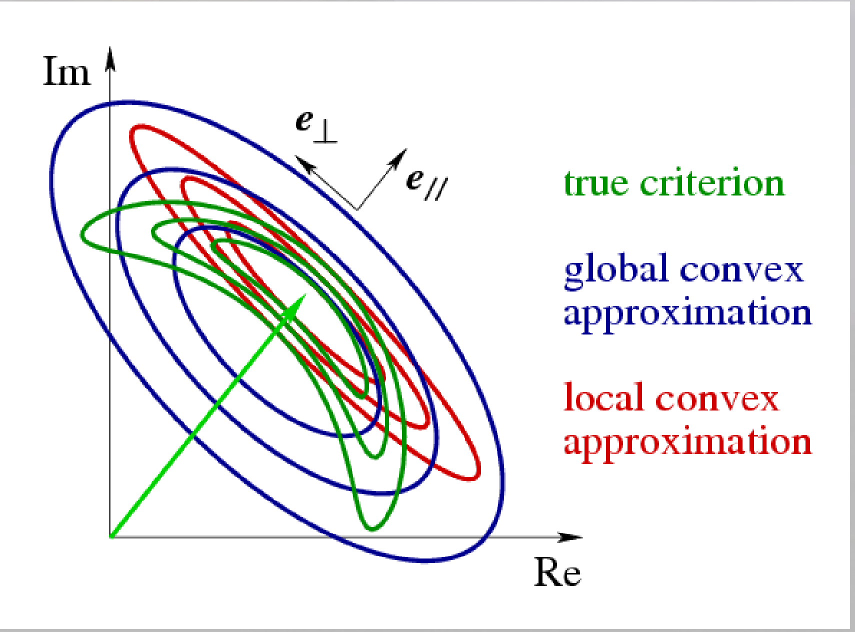
C. Hummel et al.: `http://www.mrao.cam.ac.uk/~jsy1001/exchange/complex/complex.html`

Statistics of complex optical interferometry data

for complex optical interferometric data: $\mathrm{Cov}(\rho,\phi) \simeq 0$

$$\chi^2 = \begin{pmatrix}\delta\rho\\ \delta\phi\end{pmatrix}^{\mathrm{T}} \cdot \begin{pmatrix}\sigma_\rho^2 & C_{\rho,\phi}\\ C_{\rho,\phi} & \sigma_\phi^2\end{pmatrix}^{-1} \cdot \begin{pmatrix}\delta\rho\\ \delta\phi\end{pmatrix}$$

$$\simeq \frac{\delta\rho^2}{\sigma_\rho^2} + \frac{\delta\phi^2}{\sigma_\phi^2}$$



(source: S. Meimon, 2006)

Local convex approximation

Cartesian coordinates:

$$\begin{aligned} x &= \rho \cos\phi \\ y &= \rho \sin\phi \end{aligned} \quad \Rightarrow \quad \begin{aligned} \delta x &\simeq \cos\phi\ \delta\rho - \rho \sin\phi\ \delta\phi \\ \delta y &\simeq \sin\phi\ \delta\rho + \rho \cos\phi\ \delta\phi \end{aligned}$$

$$\mathrm{Cov}(\rho, \phi) \simeq 0$$

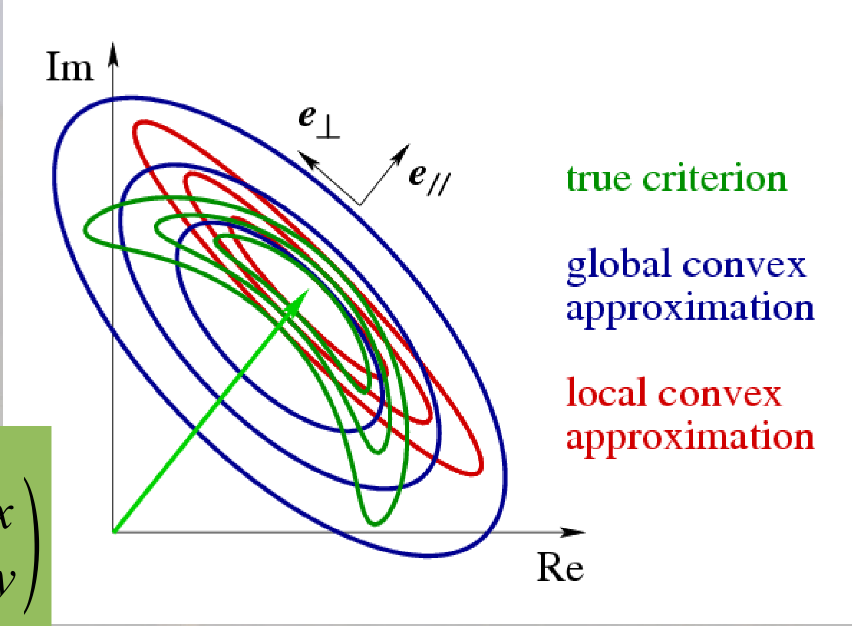
$$\begin{aligned} \sigma_x^2 &\simeq \cos^2\phi\ \sigma_\rho^2 + \rho^2 \sin^2\phi\ \sigma_\phi^2 \\ \sigma_y^2 &\simeq \sin^2\phi\ \sigma_\rho^2 + \rho^2 \cos^2\phi\ \sigma_\phi^2 \\ C_{x,y} &\simeq \sin\phi \cos\phi \left(\sigma_\rho^2 - \rho^2\ \sigma_\phi^2\right) \end{aligned} \quad \Rightarrow \quad \sigma_x^2\ \sigma_y^2 - C_{x,y}^2 \simeq \rho^2\ \sigma_\rho^2\ \sigma_\phi^2$$

$$\chi^2 = \frac{1}{\rho^2\ \sigma_\rho^2\ \sigma_\phi^2} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}^{\mathrm{T}} \cdot \begin{pmatrix} \sigma_y^2 & -C_{x,y} \\ -C_{x,y} & \sigma_x^2 \end{pmatrix} \cdot \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}$$

Statistics of real interferometric data

true **criterion:**

$$\chi^2 \simeq \frac{\delta\rho^2}{\sigma_\rho^2} + \frac{\delta\phi^2}{\sigma_\phi^2}$$



(source: S. Meimon, 2006)

convex approximation:

$$\chi^2 \simeq \frac{1}{\rho^2\,\sigma_\rho^2\,\sigma_\phi^2}\begin{pmatrix}\delta x\\ \delta y\end{pmatrix}^{\mathrm{T}}\cdot\begin{pmatrix}\sigma_y^2 & -C_{x,y}\\ -C_{x,y} & \sigma_x^2\end{pmatrix}\cdot\begin{pmatrix}\delta x\\ \delta y\end{pmatrix}$$

circular approximation (Goodman approximation):

$$\sigma_\phi = \rho\,\sigma_\rho$$
$$\Rightarrow\ \sigma_x = \sigma_y = 0 \quad \text{and} \quad \mathrm{Cov}(x,y) = 0$$

Statistics summary

log-likelihood of complex visibility:

- heterogeneous data (VIS, VIS2, T3) yields sum of chi-square terms
- complex chi-square in polar coordinates is not convex w.r.t. complex visibility (hence *approximations* below)
- Goodman approximation (circular, may be OK for low SNR's) $\sigma_\phi = \rho\,\sigma_\rho$

criterion:

- local approximation (obtained from a local expansion)
- global approximation (same moments as true criterion)
- *true* criterion
  - amplitude and phase are independent and Gaussian $\mathrm{Cov}(\rho,\phi) \simeq 0$
  - however homogeneous distribution ($\propto\rho$) is not constant in polar coordinates hence maximum likelihood yield a solution which depends on the change of variables (Tarantola, 1987)

Is the model reliable?

– chi-square statistics yield level of confidence (*i.e.* what is the probability to have found the correct model with such a bad chi-squared value);

– however this statistics is ***very sharp*** (~ Gaussian for large N_{free}):

$$\langle \chi^2 \rangle = N_{\text{free}} = N_{\text{data}} - N_{\text{param}}$$
$$\text{Var}(\chi^2) = 2\, N_{\text{free}}$$

– noise level and modelization errors must not be underestimated;

– ***in practice***, chi-square statistics cannot be used to accept or rule out a fitted model;

– however can be used to compare two models: $\dfrac{\chi^2(\boldsymbol{m}_1)}{N_1}$ vs. $\dfrac{\chi^2(\boldsymbol{m}_2)}{N_2}$

Statistics of the parameters: linear model

linear model: $\boldsymbol{m}(\boldsymbol{x}) = \boldsymbol{A}\cdot\boldsymbol{x}$

$$\boldsymbol{r} = \pm\left(\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{x})\right)$$

$$\Rightarrow \boldsymbol{C}_{\boldsymbol{r}} = \langle(\boldsymbol{r}-\bar{\boldsymbol{r}})\cdot(\boldsymbol{r}-\bar{\boldsymbol{r}})^{\mathrm{T}}\rangle = \boldsymbol{A}\cdot\langle(\boldsymbol{x}-\bar{\boldsymbol{x}})\cdot(\boldsymbol{x}-\bar{\boldsymbol{x}})^{\mathrm{T}}\rangle\cdot\boldsymbol{A}^{\mathrm{T}}$$

$$= \boldsymbol{A}\cdot\boldsymbol{C}_{\boldsymbol{x}}\cdot\boldsymbol{A}^{\mathrm{T}}$$

$$\Rightarrow \boldsymbol{C}_{\boldsymbol{x}} = \left(\boldsymbol{A}^{\mathrm{T}}\cdot\boldsymbol{C}_{\boldsymbol{r}}^{-1}\cdot\boldsymbol{A}\right)^{-1}$$

covariance matrix of parameters

correlation matrix: $\Gamma_{j,k} = \dfrac{C_{j,k}}{\sigma_j\,\sigma_k}$

Statistics of the parameters: non-linear model

non-linear model: $\boldsymbol{m}(\boldsymbol{x}) \neq \boldsymbol{A}\cdot\boldsymbol{x}$

$$\boldsymbol{m}(\boldsymbol{x}) \simeq \boldsymbol{m}(\boldsymbol{x}_{\text{best}}) + \frac{\partial \boldsymbol{m}}{\partial \boldsymbol{x}}(\boldsymbol{x}_{\text{best}})\cdot(\boldsymbol{x} - \boldsymbol{x}_{\text{best}})$$

$$\Rightarrow \boldsymbol{C}_{\boldsymbol{x}} \simeq \left(\boldsymbol{A}^{\mathrm{T}}\cdot\boldsymbol{C}_{\boldsymbol{r}}^{-1}\cdot\boldsymbol{A}\right)^{-1}$$

with $\boldsymbol{A} = \frac{\partial \boldsymbol{m}}{\partial \boldsymbol{x}}(\boldsymbol{x}_{\text{best}})$, i.e. $A_{j,k} = \frac{\partial m_j}{\partial x_k}(\boldsymbol{x}_{\text{best}})$

however depends on data error bars,
workaround if true error bars known up to a scaling parameter:

$$\boldsymbol{C}_{\boldsymbol{x}} \simeq \frac{N_{\text{free}}}{\chi^2(\boldsymbol{x}_{\text{best}})}\left(\boldsymbol{A}^{\mathrm{T}}\cdot\boldsymbol{C}_{\boldsymbol{r}}^{-1}\cdot\boldsymbol{A}\right)^{-1}$$

Non-Gaussian Statistics

reasons to not use Gaussian statistics:

- the residuals are not Gaussian (*e.g.*, Poisson noise, however central-limit theorem);
- there are outliers (bad data);
- optical interferometry data (complex valued, not a linear space);

non-Gaussian statistics:

- accounts for non-Gaussian noise and model error
- yields ***non-quadratic penalty*** (w.r.t. residuals)
- can be used to ***rule out outliers*** (*e.g.*, ℓ_2-ℓ_1 norms)

Unique Solution?

complex visibility model for a binary star:

$$m(u,v) = \alpha + (1-\alpha)\,\mathrm{e}^{-2\,\mathrm{i}\,\pi\,(u\,x + v\,y)}$$

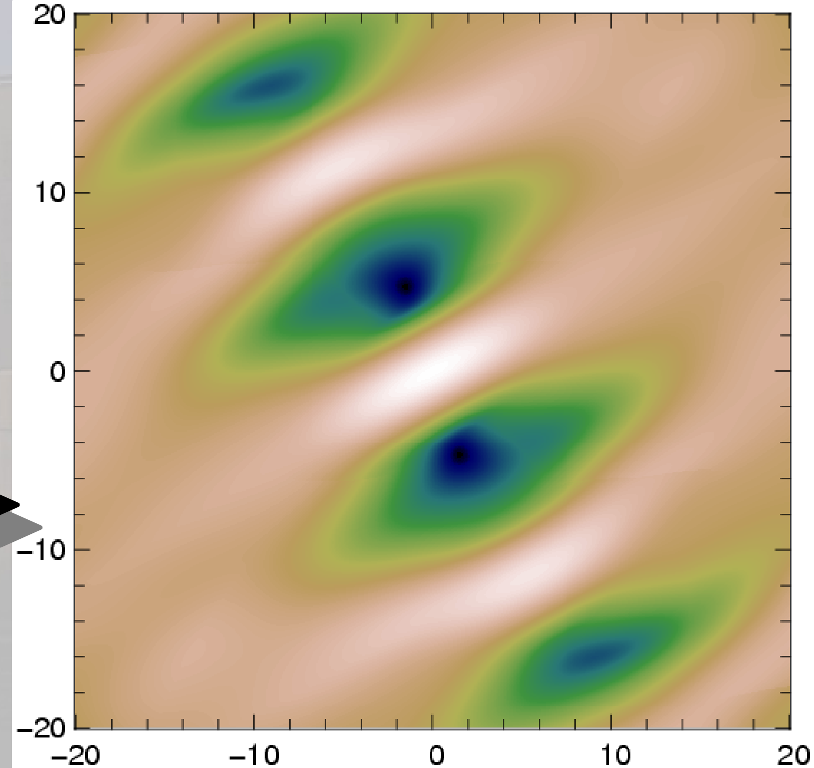
χ^2 depends on 3 parameters: α, x, and y

$$\alpha^{+}(x,y) = \arg\min_{\alpha} \chi^2(\alpha,x,y)$$

$$\chi^{2+}(x,y) = \chi^2\left(\alpha^{+}(x,y),x,y\right)$$

non convex criterion:

- ***many local minima***
- ***global optimization*** required



(χ^{2+} map, source: C. Béchet et al., 2005)

Optimization Issues

- **non convex criterion** (unless Gaussian statistics and linear model)
- many local minima
- **global optimization** required
  - systematic exploration of the parameter space
    - griding (very expensive)
    - random initial solution, then ***local optimization***
  - Monte-Carlo exploration
    - simulated annealing (*e.g.* ASA, Ingber, 1989)
    - genetic algorithms
- reduce number of parameters
  - some parameters can be uniquely estimated given the other (*e.g.* α in the binary star example if the complex visibilities are available)
- **local optimization** can however improve a given set of parameters

Local optimization: Newton method

- can be used to refine a solution
- based on Newton method:

$$f(\boldsymbol{x}+\delta\boldsymbol{x}) = f(\boldsymbol{x}) + \delta\boldsymbol{x}^{\mathrm{T}}\cdot\boldsymbol{g}(\boldsymbol{x}) + \frac{1}{2}\,\delta\boldsymbol{x}^{\mathrm{T}}\cdot\boldsymbol{H}(\boldsymbol{x})\cdot\delta\boldsymbol{x} + o\left(\|\delta\boldsymbol{x}\|^2\right)$$

where

$$\boldsymbol{g}(\boldsymbol{x}) \equiv \nabla f(\boldsymbol{x}) \qquad g_j(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial x_j}$$

$$\boldsymbol{H}(\boldsymbol{x}) \equiv \nabla\nabla f(\boldsymbol{x}) \qquad H_{j,k}(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial x_j\,\partial x_k}$$

local quadratic approximation:

$$f(\boldsymbol{x}+\delta\boldsymbol{x}) - f(\boldsymbol{x}) \simeq q(\delta\boldsymbol{x}) \equiv \delta\boldsymbol{x}^{\mathrm{T}}\cdot\boldsymbol{g}(\boldsymbol{x}) + \frac{1}{2}\,\delta\boldsymbol{x}^{\mathrm{T}}\cdot\boldsymbol{H}(\boldsymbol{x})\cdot\delta\boldsymbol{x}$$

optimal step:

$$\delta\boldsymbol{x}_{\mathrm{quad}} = \arg\min_{\delta\boldsymbol{x}} q(\delta\boldsymbol{x}) = -\boldsymbol{H}(\boldsymbol{x})^{-1}\cdot\boldsymbol{g}(\boldsymbol{x})$$

Local optimization: avoiding too large steps

Newton step can be too large (outside region where quadratic approximation is valid)

solve the constrained problem:

$$\delta \boldsymbol{x}_{\mathrm{TR}} = \arg\min_{\delta \boldsymbol{x}} q(\delta \boldsymbol{x}) \quad \text{s.t.} \quad \|\delta \boldsymbol{x}\| \leqslant \Delta$$

size of the trust region

metric:

$$\|\delta \boldsymbol{x}\| = \sqrt{\delta \boldsymbol{x}^{\mathrm{T}} \cdot \boldsymbol{D} \cdot \delta \boldsymbol{x}}$$

Lagrangian:

$$L(\delta \boldsymbol{x}, \lambda) = q(\delta \boldsymbol{x}) + \frac{1}{2} \lambda \|\delta \boldsymbol{x}\|^2$$

constrained step:

$$\delta \boldsymbol{x}_{\mathrm{TR}} = \delta \boldsymbol{x}_{\lambda} = \arg\min_{\delta \boldsymbol{x}} L(\delta \boldsymbol{x}, \lambda) = -\left(\boldsymbol{H}(\boldsymbol{x}) + \lambda\, \boldsymbol{D}\right)^{-1} \cdot \boldsymbol{g}(\boldsymbol{x})$$

Trust Region Algorithm

The algorithm is as follows (Moré & Sorensen, 1983):

0. choose an initial trust region radius Δ

1. find Lagrange multiplier λ such that:

$$\begin{array}{llll} \text{either} & \lambda = 0 & \text{and} & \|\delta \boldsymbol{x}_\lambda\| < \Delta \\ \text{or} & \lambda > 0 & \text{and} & \|\delta \boldsymbol{x}_\lambda\| \simeq \Delta \end{array}$$

2. compute goodness of quadratic approximation: $\eta = \dfrac{f(\boldsymbol{x}+\delta \boldsymbol{x}_\lambda) - f(\boldsymbol{x})}{q(\delta \boldsymbol{x}_\lambda)}$

- reject the step $\delta \boldsymbol{x}$ if η too small
- enlarge trust region radius Δ if $\eta \sim 1$, reduce Δ if η too small

3. check for convergence of repeat with step 1

Levenberg-Marquardt algorithm (1)

local minimization of a sum of squares criterion

$$f(\boldsymbol{x}) = \sum_{j=1}^{N_{\text{data}}} e_j^2(\boldsymbol{x})$$

e.g. Gaussian independent noise:

$$\Pr(\boldsymbol{r}) = \prod_{j}^{N_{\text{data}}} \Pr(r_j) = \prod_{j}^{N_{\text{data}}} \frac{\exp\left(\frac{-r_j^2}{2\,\sigma_j^2}\right)}{\sqrt{2\,\pi\,\sigma_j}}$$

the *e*'s are normalized residuals errors:

$$e_j(\boldsymbol{x}) = \frac{r_j(\boldsymbol{x})}{\sigma_j} = \pm\frac{d_j - m_j(\boldsymbol{x})}{\sigma_j}$$

Levenberg-Marquardt algorithm (2)

criterion:

$$f(\boldsymbol{x}) = \sum_{j=1}^{N_{\text{data}}} e_j^2(\boldsymbol{x})$$

gradient and Hessian:

$$g_k(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial x_k} = 2 \sum_{j=1}^{N_{\text{data}}} \frac{\partial e_j(\boldsymbol{x})}{\partial x_k} \, e_j(\boldsymbol{x})$$

$$H_{k,l}(\boldsymbol{x}) = \frac{\partial^2 f(\boldsymbol{x})}{\partial x_k \, \partial x_l} = 2 \sum_{j=1}^{N_{\text{data}}} \frac{\partial e_j(\boldsymbol{x})}{\partial x_k} \, \frac{\partial e_j(\boldsymbol{x})}{\partial x_l} + 2 \sum_{j=1}^{N_{\text{data}}} \frac{\partial^2 e_j(\boldsymbol{x})}{\partial x_k \, \partial x_l} \, e_j(\boldsymbol{x})$$

$$\simeq 2 \sum_{j=1}^{N_{\text{data}}} \frac{\partial e_j(\boldsymbol{x})}{\partial x_k} \, \frac{\partial e_j(\boldsymbol{x})}{\partial x_l}$$

hence, only 1st order partial derivatives needed:

$$J_{j,k} \equiv \frac{\partial e_j(\boldsymbol{x})}{\partial x_k}$$

Levenberg-Marquardt algorithm summary

criterion (sum of squares):

$$f(\boldsymbol{x}) = \sum_{j=1}^{N_{\text{data}}} e_j^2(\boldsymbol{x})$$

iteratively minimized by a ***trust region method***

quadratic approximation:

$$f(\boldsymbol{x}+\delta\boldsymbol{x}) - f(\boldsymbol{x}) \simeq q(\delta\boldsymbol{x}) \equiv \delta\boldsymbol{x}^{\mathrm{T}}\cdot\boldsymbol{g}(\boldsymbol{x}) + \frac{1}{2}\,\delta\boldsymbol{x}^{\mathrm{T}}\cdot\boldsymbol{H}(\boldsymbol{x})\cdot\delta\boldsymbol{x}$$

with

$$\boldsymbol{g}(\boldsymbol{x}) = 2\,\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\cdot\boldsymbol{e}(\boldsymbol{x})$$
$$\boldsymbol{H}(\boldsymbol{x}) \simeq 2\,\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\cdot\boldsymbol{J}(\boldsymbol{x})$$

and

$$J_{j,k}(\boldsymbol{x}) \equiv \frac{\partial e_j(\boldsymbol{x})}{\partial x_k}$$

(can be obtained by finite differences)

Closure (*i.e.* lunch time!)

model fitting:

- gives you the *best* model parameters and their error bars providing
  - your model is pertinent
  - the statistics of the errors is not too far from Gaussian
  - you have found the global minimum
  - the statistics may be truly multi-modal (*i.e.* the other local minima may deserve some attention)
- ***is real data processing*** (not a magic black box) you have to understand what is undergone

issues not addressed in this talk:

- global optimization (non-convex criterion)
- estimation of partial derivatives by finite differences
- accounting for correlations in the data (however see general Gaussian)
- ***residual definitions for non-Gaussian data (phase, amplitude)***

Some References

mathematics & statistics:

- A. Tarantola, *Inverse Problem Theory - Methods for Data Fitting and Model Parameter Estimation*, Elsevier, 1987 (re-edited by SIAM).
- D. Pelat, Bruits et signaux, cours de l'école doctorale Astrophysique et Techniques Spatiales (Univ. Paris 6 & 7), 1998.

optimization:

- J.J. Moré & D.C. Sorensen, *Computing A Trust Region Step*, SIAM J. Sci. Stat. Comp. **4**, 553-572, 1983.
- W.H., Press, S.A. Teukolsky, W.T. Vetterling & B.P. Flannery, *Numerical Recipes in C*, Cambridge University Press, 1992.
- L. Ingber, *Very fast simulated re-annealing*, Math. Comput. Modelling **12**, 967-973, 1989.

optical interferometry:

- T.A. Pauls, J.S. Young, W.D. Cotton & J.D. Monnier, *A Data Exchange Standard for Optical (Visible/IR) Interferometry*, PASP **117**, 1255-1262, 2005.
- S. Meimon, L.M. Mugnier & G. le Besnerais, *Convex approximation to the likelihood criterion for aperture synthesis imaging*, J. Opt. Soc. Am. A **22**, 2348-2356, 2005.
- M. Tallon, I. Tallon-Bosc, E. Thiébaut, C. Béchet, M. Chadid, A. Chelli, J.-M. Clausse, G. Duvert, D. Mourard, G. Perrin, R. Petrov, Ph. Stee, G. Zins, *JMMC Model Fitting Software for optical interferometry*, SF2A Conference Series, p. 171-174, 2004.
- C. Béchet, M. Tallon, I. Tallon-Bosc, E. Thiébaut, *Model Fitting Software developed by JMMC for optical interferometry data analysis*, SF2A Conference Series, p. 271-274, 2005.