Characterising the Information Content of Microlensing Light Curves

Markus Hundertmark

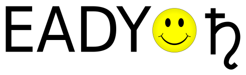
University of St Andrews

Feb. 16th, 2012

Shannon's definition of information content

$$I_S = -\log_2(p(x)) \Rightarrow H = \langle -\log_2(p(x)) \rangle$$

Frequently observed values are less informative.



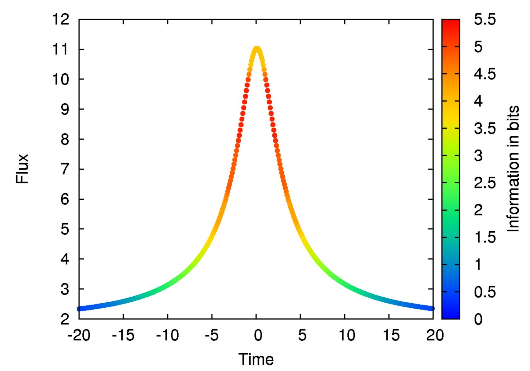
$\Rightarrow$ y cn rd txts wtht vwls!

Application

Efficient data transfer: e.g. Morse code:

E: · Q: — — ·— 9: — — — — ·

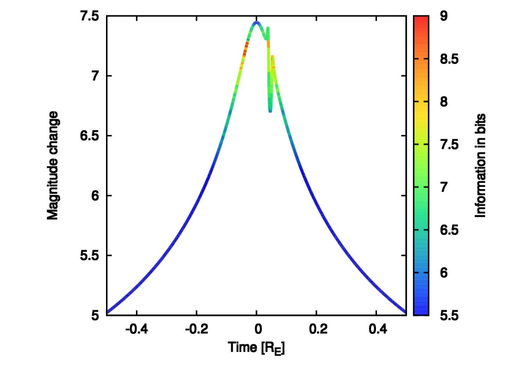
Shannon's definition of information



PSPL light curve and its information content

- Characters are replaced by bins (finite accuracy)
- The information content (with respect to μ) follows the gradient
- High-magnification events are preferred

Low mass-ratio event



Features are not included

- Features seem to be less valuable (counterintuitive)
- Scheme for binning data
- A parameter dependent measure of information is required

From Shannon's information to the Fisher matrix

Light curves are characterised based on maximum likelihood estimators $\mathscr{L}$. Information can be understood as relative entropy:

Relative entropy (KL-divergence)

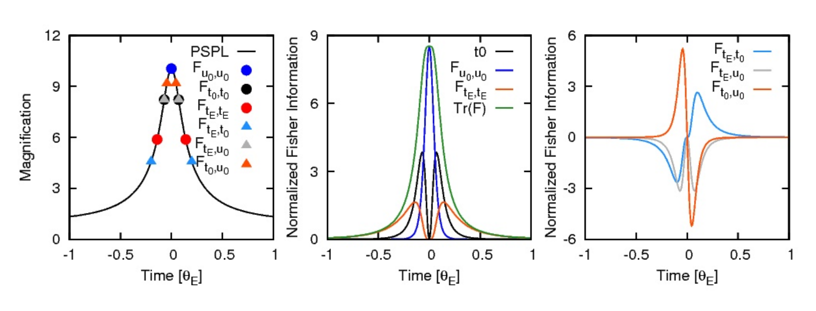
$$\langle \log(\mathscr{L}(p)) - \log(\mathscr{L}(p+\Delta p)) \rangle$$

Expanding this expression leads to the Fisher matrix - a metric measure of the manifold of models:

Fisher's interpretation

$$I_{\mathrm{F}} = \left\langle \left(\frac{\partial \log(\mathscr{L})}{\partial p_i} \right) \left(\frac{\partial \log(\mathscr{L})}{\partial p_j} \right) \right\rangle$$

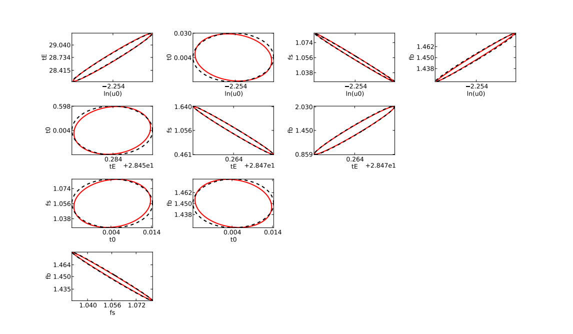
Fisher Information in light curves



Not all points contribute equally to our knowledge of parameters

- The Fisher matrix is independent of a given realisation.
- *First maximum (50 % blending):* $\Delta I_{\mathbf{mag}} > 0.3$

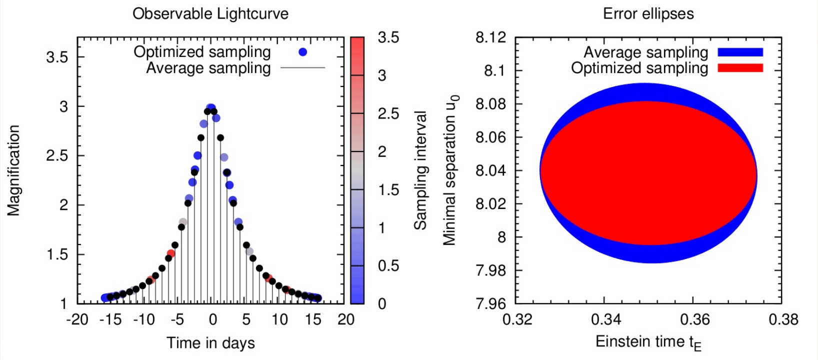
Fisher Matrix vs. MCMC covariance matrices



MCMC: dashed black, Fisher matrix: red

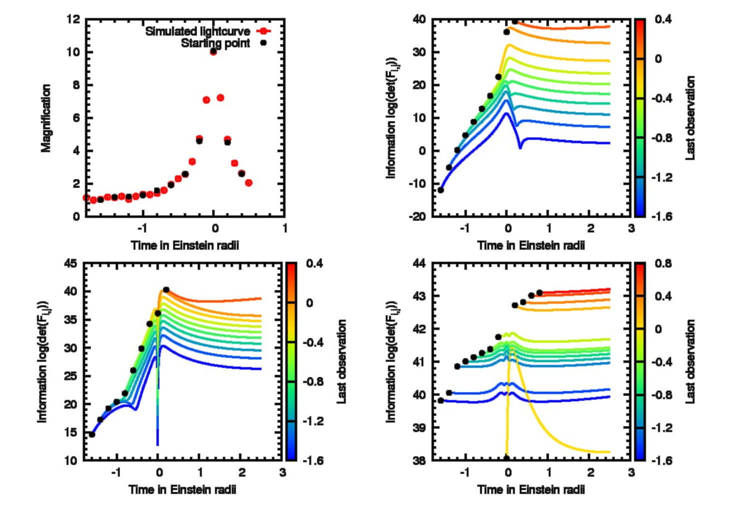
- MCMC and Fisher matrix agree for: appropriately scaled uncertainties, unbiased estimators and Gaussian noise
- linearize model (multivariate Gaussian)

Information modulated strategy

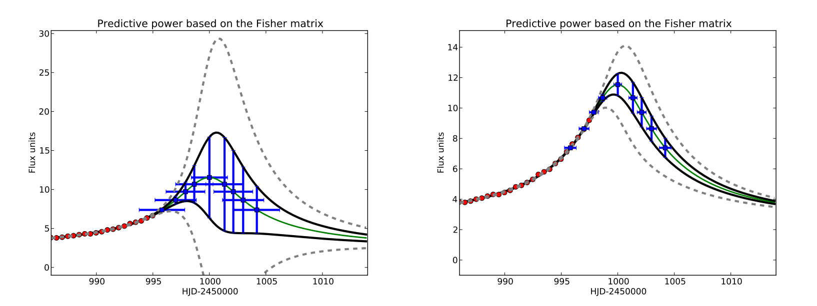


- More information from the same observing time
- Event parameters need to be known in advance...

Constructive strategy



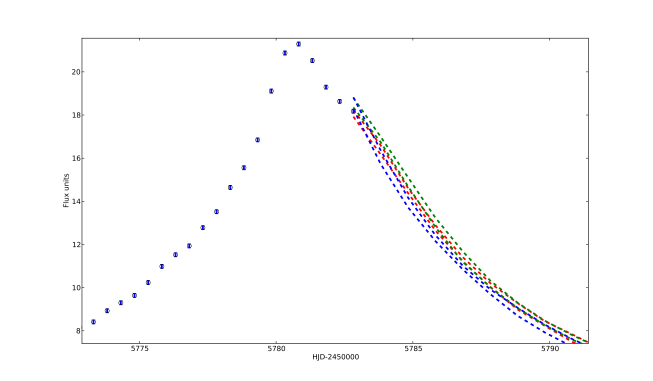
Application: Anomaly detection



Prediction for a given observation, model and likelihood

- The Fisher matrix provides an estimate of the maximal predictive power
- $\Rightarrow$ Pre-peak anomaly detection and sampling suggestion

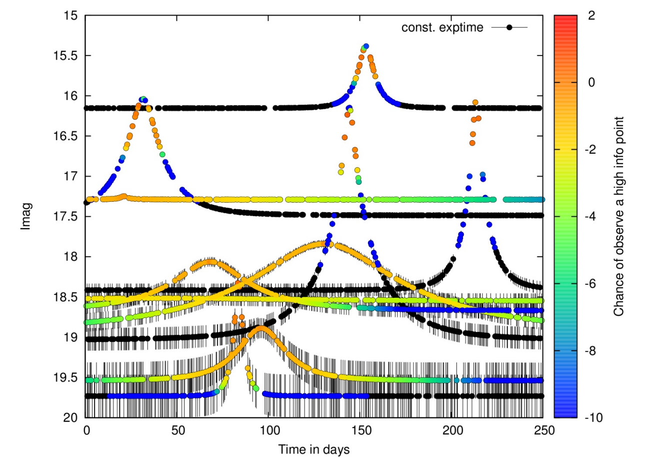
Application: Binary model discrimination



Fisher matrix estimates based on fully automated binary fits (see talk by Bozza)

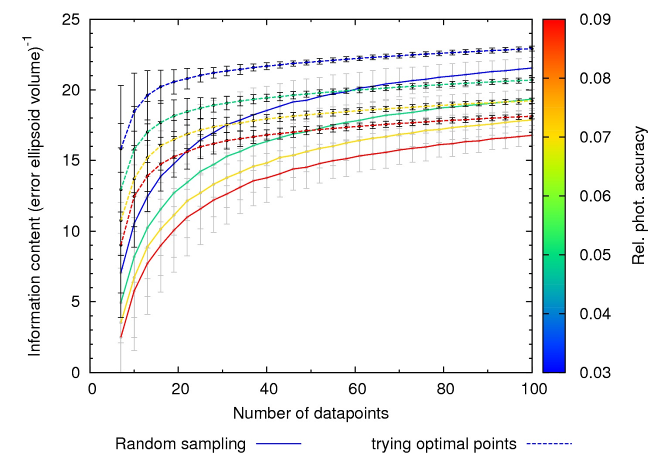
- Discriminate between different binary models
- Detecting anomalous deviations (e.g. triple lens)

Refining observing strategies



Strategy: Try to include information carrying points

Improvement

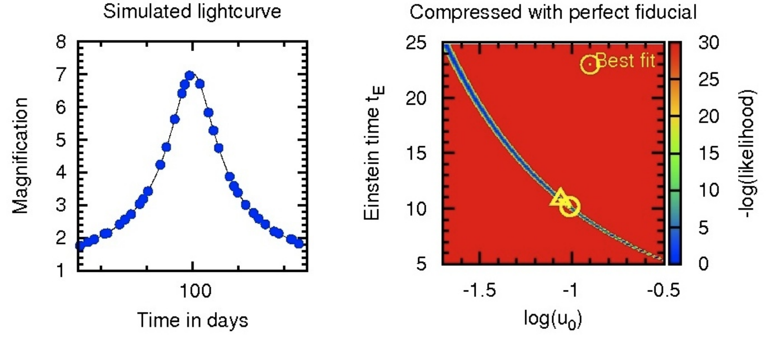


- Expected information always better
- Typical improvement 2-3 orders or magnitude

Massive "lossless" compression MOPED [1]

- Basic idea: weighted measurements
- compress the measurement vector to a single value

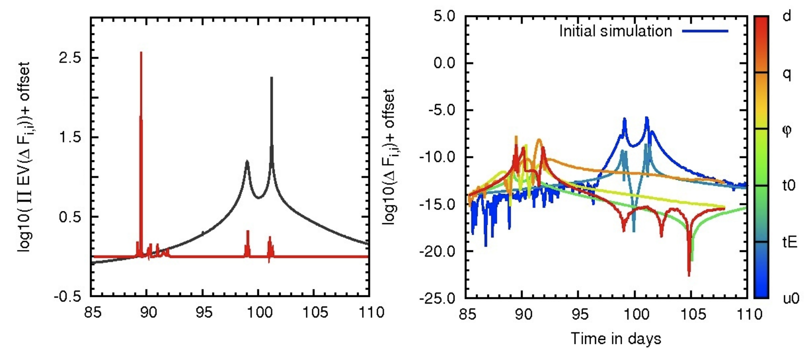
$\Rightarrow n_{D,compressed} = n_{parameters}$



The Fisher matrix is conserved, but the estimate is biased

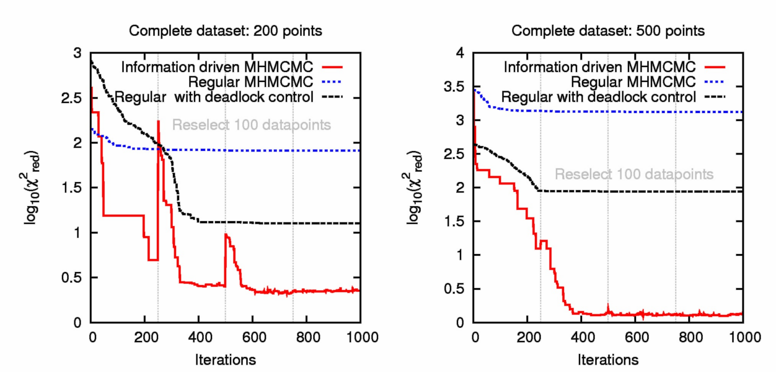
[1] Heavens, A. F., Jimenez, R., & Lahav, O. 2000

Information content in binary light curves



- Identify information carrying data-points (model dependent)
- Advantage: locally adaptive (gaps at irrelevant parts)

Information driven MCMC



- Reduced computation time and faster descend to lower $ln(L)$
- NB: Sampling from converged MCs requires the full dataset

Conclusions and future work

- Optimal experimental design offers a formalism for increasing our knowledge for a given number of observations
- Observations can be placed for improving the characterisation of events
- The information content offers ways of compressing datasets before the analysis
- The calculation is easy, reproducible **but** can be biased.
- Data compression:
  - Accelerate the analysis of binary events
  - Excludes less likely models

Thank you for your attention!