Strong Lens Modeling (III): Advanced Techniques

Chuck Keeton

Rutgers, the State University of New Jersey

Goals

point sources + parametric lens models

- composite models
- astrophysical priors
- substructure
- statistical techniques

extended sources

free-form lens models

Case studies

What can you do with advanced analyses of point sources and parametric lens models?

Q0957+561

- sophisticated composite models
- use of astrophysical priors
- MCMC

HE 0435−1223

- substructure
- statistical methods
- nested sampling

(work led by Ross Fadely)

Q0957+561

Fadely et al. 2010ApJ...711..246F

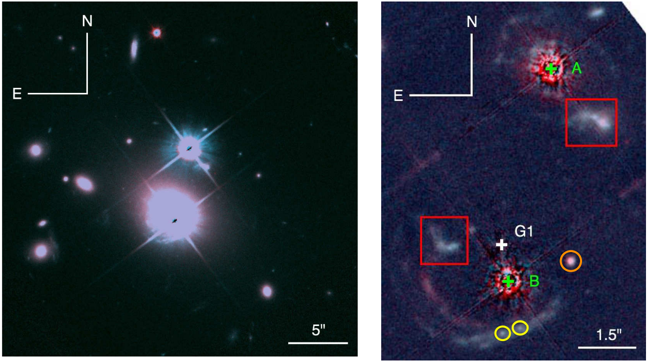


Figure: *(Left)* Central $30''$ of combined HST F606W and F814W images. *(Right)* Close-up of the strong lensing region, after the main lensing galaxy and quasar images have been subtracted.

Lensed features

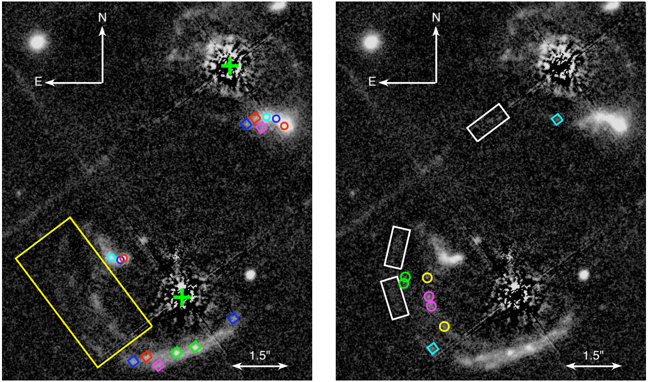


Figure: Sets of multiple images — all told, 30 images of 14 sources.

Properties of main lens galaxy

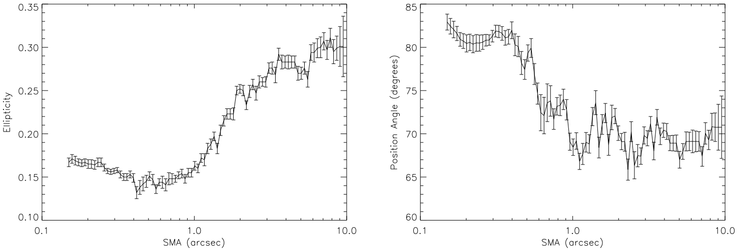


Figure: Ellipticity and position angle of galaxy isophotes.

Components of mass model

Stellar component: observed light distribution

- stellar mass-to-light ratio Υ

Dark matter halo: NFW or softened power law

- normalization
- scale radius
- ellipticity and position angle

Environment: cluster surrounding main lens galaxy

$$\begin{aligned}\phi_{\rm env}(r,\theta) &= \frac{\kappa_c}{2}\,r^2 + \frac{\gamma}{2}\,r^2\cos 2(\theta-\theta_\gamma) \\ &\quad + \frac{\sigma}{4}\,r^3\cos(\theta-\theta_\sigma) + \frac{\delta}{6}\,r^3\cos 3(\theta-\theta_\delta) + \ldots\end{aligned}$$

- shear (γ, θ_γ), higher-order terms $(\sigma, \theta_\sigma, \delta, \theta_\delta)$
- mass sheet κ_c constrained with separate weak lensing analysis

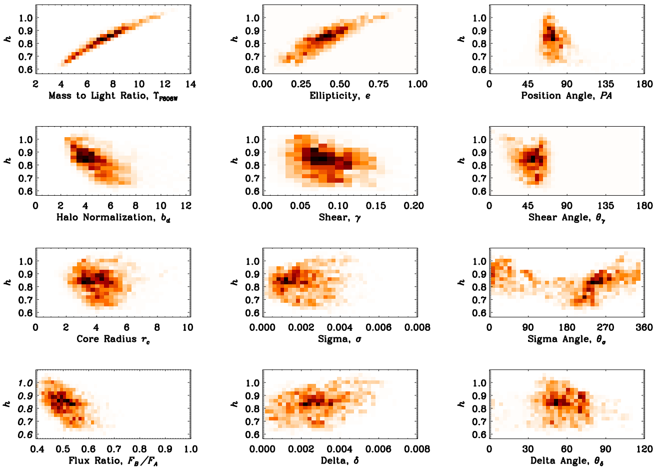
Searching parameter space

Full set of parameters:

- 11 mass model parameters — searched explicitly (MCMC)
- 28 source position parameters — optimized analytically
- H_0 from time delay

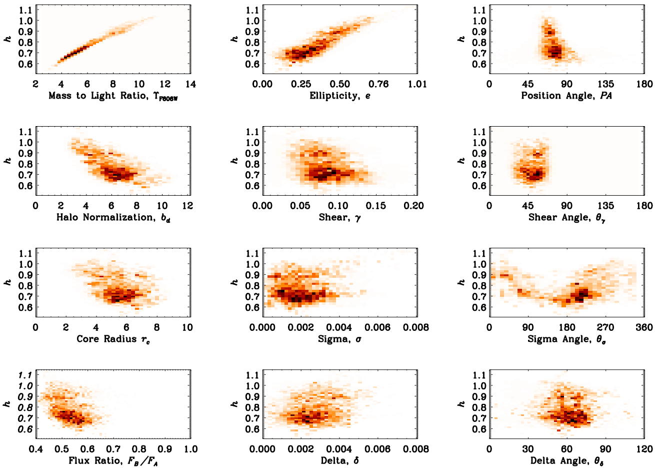
Basic results

Softened power law halo with isothermal profile ($\alpha = 1$)



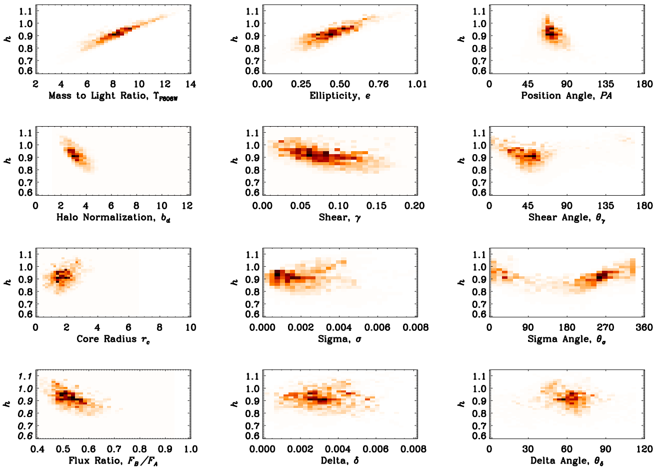
Basic results

Softened power law halo with steeper profile ($\alpha = 0.5$)



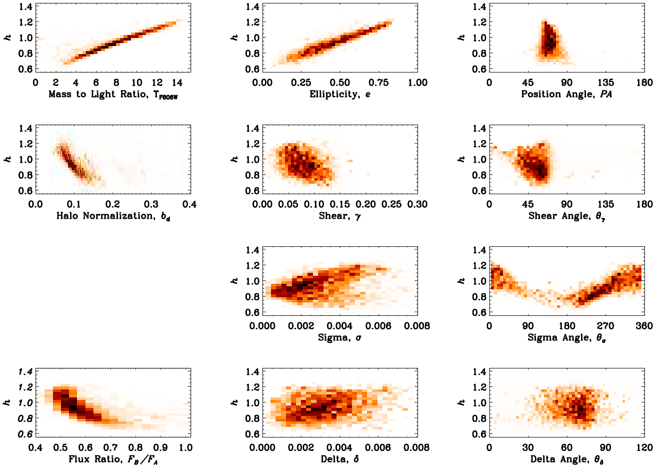
Basic results

Softened power law halo with shallower profile ($\alpha = 1.5$)



Basic results

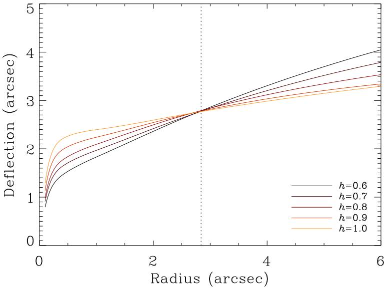
NFW halo



Density profile

Deflection curve, $\alpha(r) \propto M(r)/r$ — 2-d analog of rotation curve

Trade-off between stars and dark matter changes density profile



Rising deflection curve ⇒ density profile shallower than isothermal.
Due to massive cluster around lens?

Additional information – priors

Stellar mass-to-light ratio is not totally unknown.

Can predict it using Stellar Population Synthesis (SPS) models.

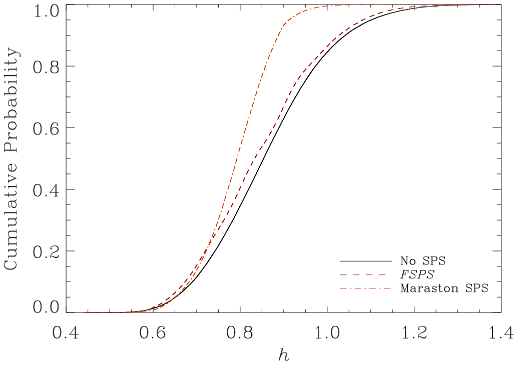
- generate a population of stars at some time
- stellar evolution models $\rightarrow$ predict how pop'n evolves
- stellar atmospheres $\rightarrow$ predict spectrum as a function of time
- include star formation history $\rightarrow$ predict galaxy spectrum

e.g., Bruzual & Charlot 2003MNRAS.344.1000B; Maraston et al. 2009MNRAS.394L.107M; Conroy et al. 2009ApJ...699..486C

Fit SPS models to observed galaxy colors, constrain Υ.

Note: analysis depends on H_0 through time vs. redshift.

Combined constraints on H_0



SPS constraints $\rightarrow H_0 = 79.3^{+6.7}_{-8.5}$ km s^{-1} Mpc^{-1} (68% CL)

Priors on H_0

Instead of trying to recover H_0, we could place priors from independent measurements.

Distance ladder (Riess et al. 2009ApJ...699..539R):

$$H_0 = 74.2 \pm 3.6 \text{ km s}^{-1} \text{ Mpc}^{-1}$$

WMAP5+SNe+BAO (Komatsu et al. 2009ApJS..180..330K):

$$H_0 = 70.5 \pm 1.3 \text{ km s}^{-1} \text{ Mpc}^{-1}$$

Stellar mass-to-light ratio

From lensing, using priors on H_0:

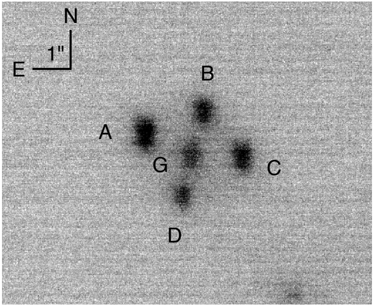
$$\Upsilon = 5.5^{+0.9}_{-0.5} \qquad \text{(distance ladder priors)}$$
$$\Upsilon = 5.5^{+0.2}_{-0.3} \qquad \text{(WMAP5+SNe+BAO priors)}$$

SPS models:

$$\Upsilon = 5.9 \pm 1.9$$

Use lensing to constrain stellar populations?!

HE 0435−1223



Constraints

- HST positions, $\sigma = 3$–5 mas
- optical/IR fluxes, $\sigma \sim 5\%$
- (time delays, $\sigma = 0.8$ d)

(Fadely & CRK 2011AJ....141..101F, 2012MNRAS.419..936F)

Statistics: Comparing models

Bayesian evidence allows objective model comparison, even with different numbers of parameters.

$$Z(M) = \int \mathcal{L}(d|\mathbf{q}, M)\ P(\mathbf{q}, M)\ d\mathbf{q}$$

Compare two models via Z_2/Z_1 or $\log_{10}(Z_2/Z_1) = \Delta\log_{10}(Z)$.

Jeffreys (1961) scale:

$\Delta\log_{10}(Z)$	Significance
0–0.5	Barely worth mentioning
0.5–1.0	Substantial
1.0–1.5	Strong
1.5–2.0	Very strong
> 2.0	Decisive

HE0435: Smooth mass models

parameters

- main galaxy: mass, position, e/PA, core radius, profile (7)
- neighbor galaxy: mass, position, e/PA (5)
- rest of environment: shear/PA (2)
- source: position, flux (3)

data

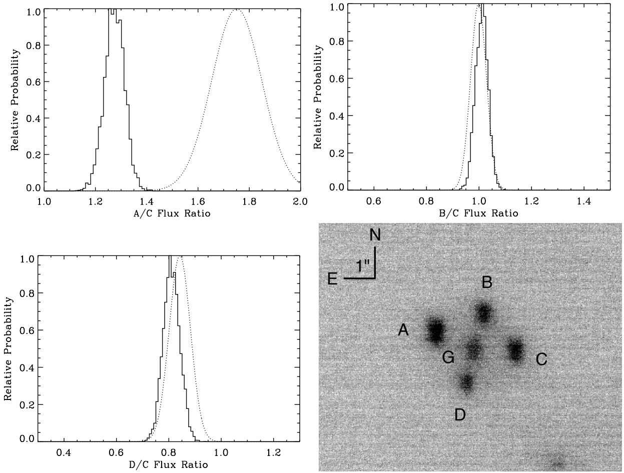
- images: positions, fluxes (12)
- main galaxy: position (2)
- neighbor galaxy: position (2)

HE0435: Smooth mass models

16 constraints, 17 parameters — but best $\chi^2 = 24.6$ (!)

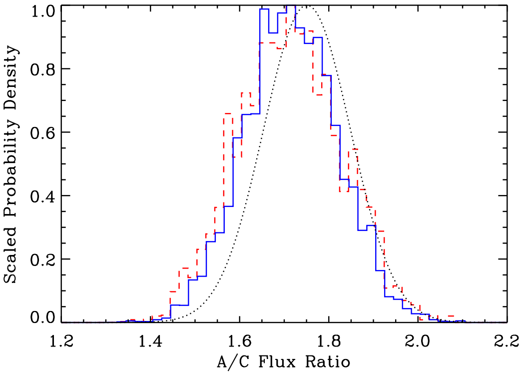
HE0435: Smooth mass models

16 constraints, 17 parameters — but best $\chi^2 = 24.6$ (!)



With mass clump(s)

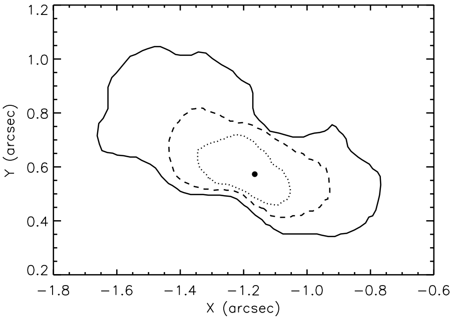
Add one clump near image A.
Add three clumps near images A, B, D.
Clumps are truncated isothermal spheres.



Position of clump A

95% confidence limits

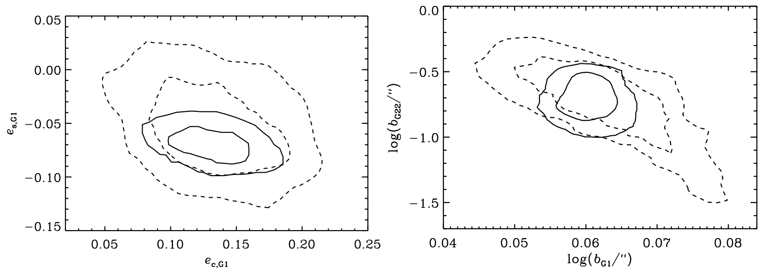
- dotted: $M < 10^6\,M_\odot$
- dashed: $M < 10^7\,M_\odot$
- solid: $M < 10^8\,M_\odot$



Posterior parameter constraints

68% and 95% confidence intervals

- solid: smooth model
- dashed: + clump A



Statistical significance of clump(s)

Use nested sampling to compute Bayesian evidence and compare different models.

model	$\Delta \log_{10}(Z)$
smooth	$\equiv 0$
clump A	3.83 ± 0.12
clumps AD	3.90 ± 0.13
clumps AB	4.46 ± 0.12
clumps ABD	4.35 ± 0.13

Decisive evidence for a clump near image A.

$$\log_{10}(M^A_{\rm ein}) = 7.65^{+0.87}_{-0.84} \qquad \log_{10}(M^A_{\rm tot}) = 9.31^{+0.44}_{-0.42}$$

Intriguing evidence for a second clump near image B.

$$\log_{10}(M^B_{\rm ein}) = 6.55^{+1.01}_{-1.51} \qquad \log_{10}(M^B_{\rm tot}) = 8.76^{+0.50}_{-0.77}$$

Full population of clumps

It seems unlikely that the lens galaxy contains one or two clumps that are (almost) perfectly aligned with the quasar images.

More likely: they are "special" representatives of a larger pop'n.

Try to constrain the population directly

- assume truncated isothermal spheres with mass function

$$\frac{dN}{dm} \propto m^{-1.9}, \qquad m \in 10^7\text{–}10^{10}\,M_\odot$$

- see whether models make sense, constrain $\kappa_s = \Sigma_s/\Sigma_{\rm crit}$

Statistical inference

Parameters

- q = smooth model
- s = substructure *population* (abundance, mass function, etc.)
- c = individual clumps (position, mass, etc.)

Most interested in marginalized posterior for substructure population parameters:

$$P(s) \propto \int \mathcal{L}(c, q)\ P(c|s)\ P(s, q)\ dc\ dq$$

Monte Carlo techniques

Need to evaluate

$$P(s) \propto \int \mathcal{L}(c,q)\ P(c|s)\ P(s,q)\ dc\ dq$$

We can't do the c integral explicitly!

Use Monte Carlo integration: let c_j be a realization of the clump population, drawn from $P(c|s)$. Then

$$P(s) \propto \sum_j \int \mathcal{L}(c_j,q)\ P(s,q)\ dq$$

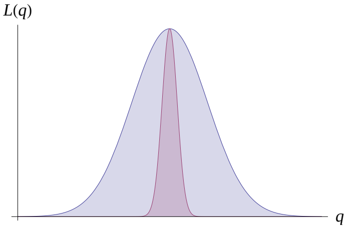
Marginalizing vs. optimizing

$$P(s) \propto \sum_j \int \mathcal{L}(c_j, q)\ P(s, q)\ dq$$

For each c_j, what do we do with q?

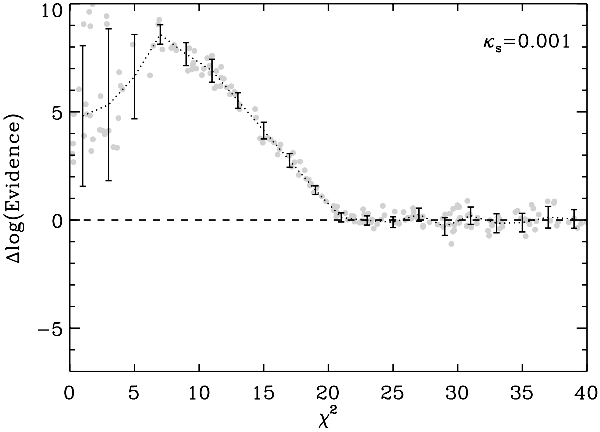
- Marginalize = do the integral, find the **area**
- Optimize = just find the **peak**

They are not necessarily equivalent!



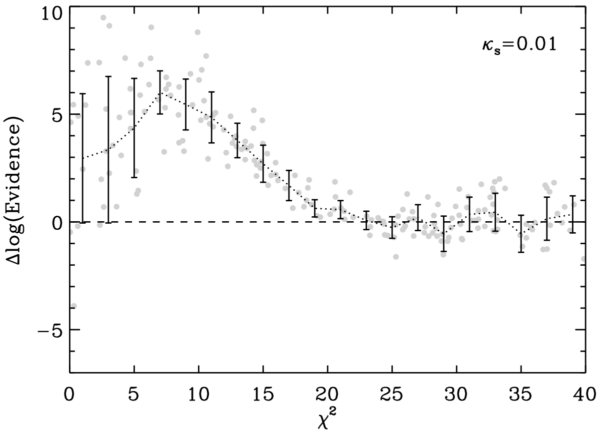
Marginalizing vs. optimizing

Each point is one realization of clump pop'n; $\mathcal{L}_{\text{peak}} = e^{-\chi^2/2}$



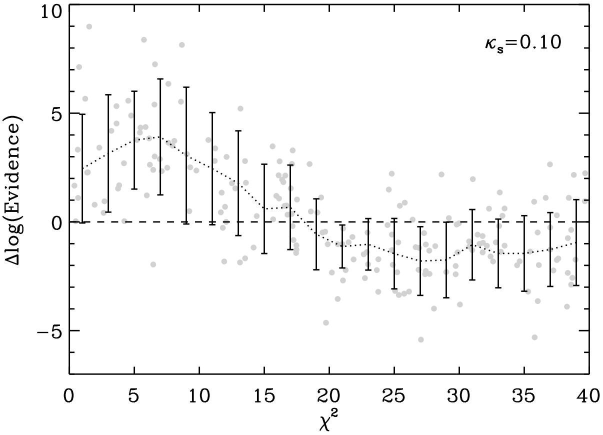
Marginalizing vs. optimizing

Each point is one realization of clump pop'n; $\mathcal{L}_{\text{peak}} = e^{-\chi^2/2}$



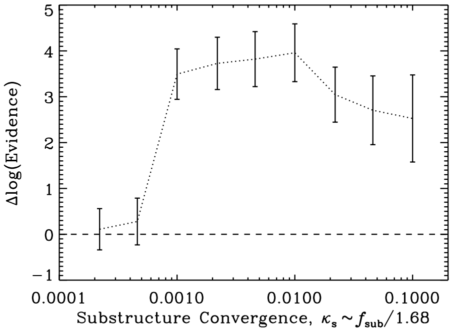
Marginalizing vs. optimizing

Each point is one realization of clump pop'n; $\mathcal{L}_{\text{peak}} = e^{-\chi^2/2}$



Results

Recall: $dN/dm \propto m^{-1.9}$ for $m \in 10^7$–$10^{10}\ M_\odot$



$\Rightarrow$ $f_{\rm sub} > 0.00077$ at Einstein radius

Extended source lenses: Arcs and rings

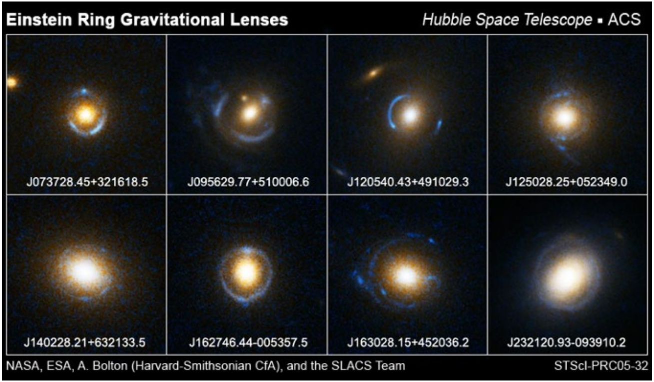


Figure: Arcs and rings from SLACS (http://www.slacs.org).

Source reconstruction

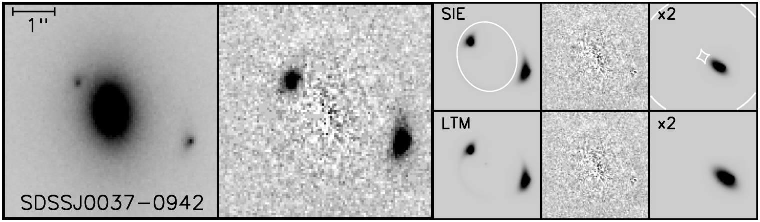


Figure: Example of source reconstruction in a SLACS lens (Bolton et al. 2008ApJ...682..964B)

Source reconstruction

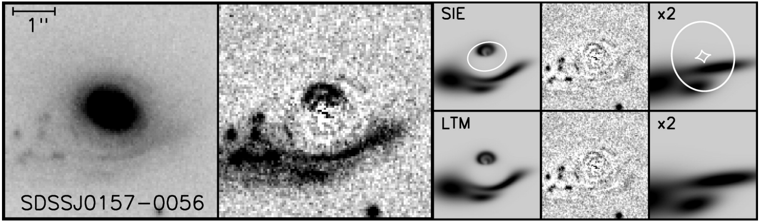
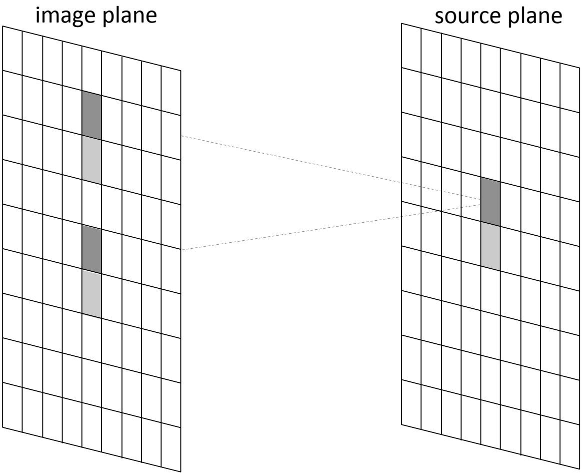


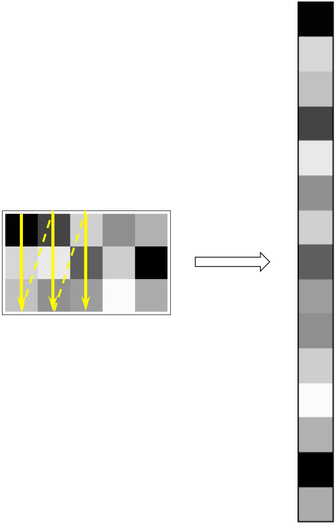
Figure: Example of source reconstruction in a SLACS lens (Bolton et al. 2008ApJ...682..964B)

Lensing conserves surface brightness

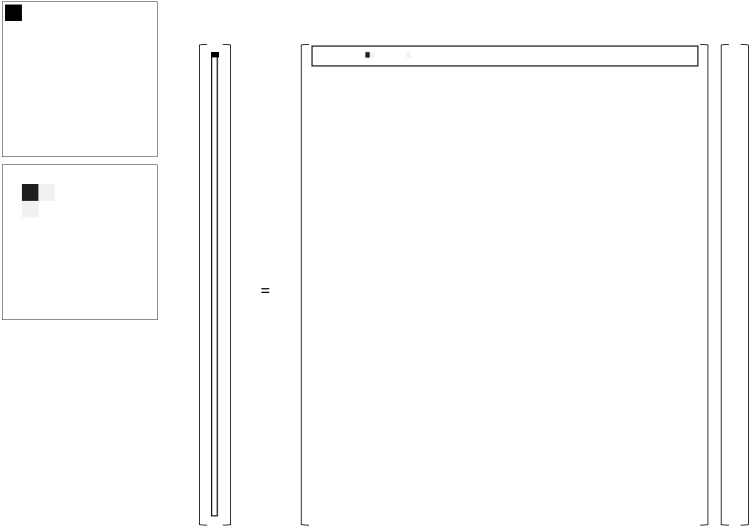


$$\mathbf{d} = \mathbf{L}\,\mathbf{s}$$

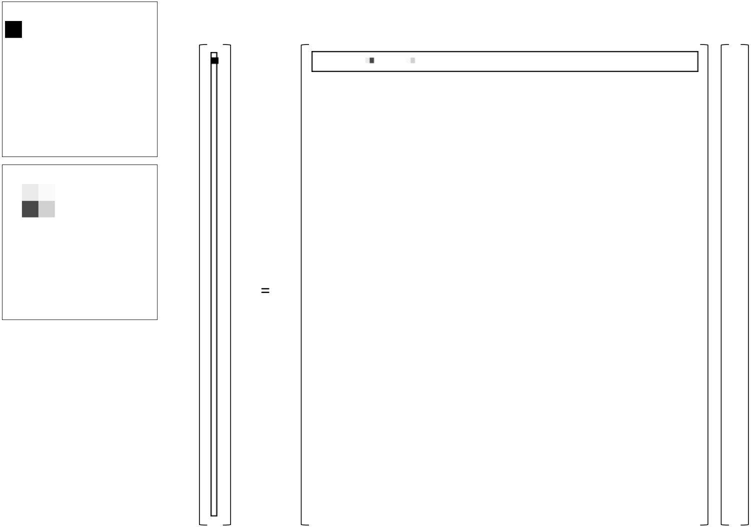
"Unfold" 2-d image into vector



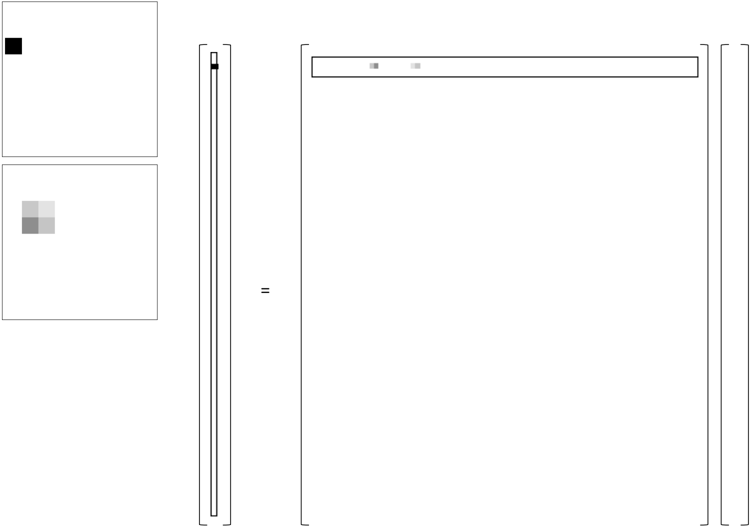
Constructing the lensing operator



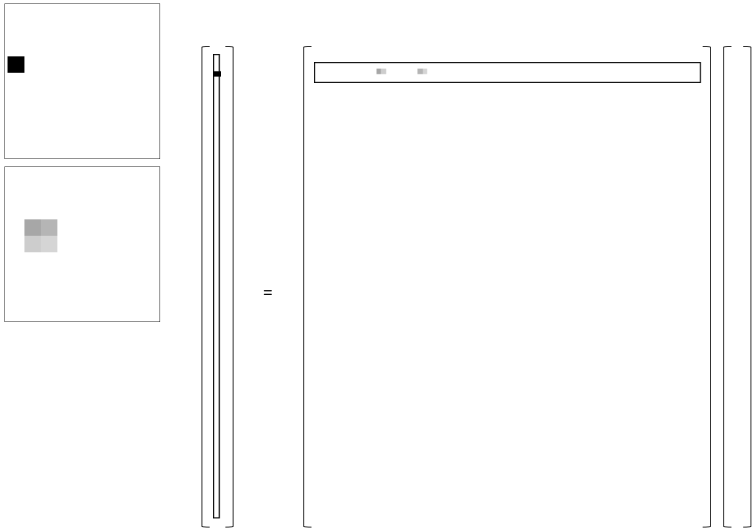
Constructing the lensing operator



Constructing the lensing operator



Constructing the lensing operator



Pixelated sources

pure surface brightness map:

$$\mathbf{d} = \mathbf{L}_0 \, \mathbf{s}$$

with PSF:

$$\mathbf{d} = \mathbf{L} \, \mathbf{s} \qquad \text{where} \quad \mathbf{L} = \mathbf{B} \, \mathbf{L}_0$$

goodness of fit:

$$\chi^2_{\rm img} = (\mathbf{L}\,\mathbf{s} - \mathbf{d}^{\rm obs})^t\,\mathbf{S}_d^{-1}\,(\mathbf{L}\,\mathbf{s} - \mathbf{d}^{\rm obs})$$

in general, more parameters than constraints, so a large family of solutions

many of the solutions may be unphysical (e.g., lots of negative flux) or merely implausible (e.g., spikes or weird shapes)

Regularization

goal: penalize models that are unrealistic

- penalize spikes

$$\chi^2_{\rm reg} \sim \sum s_j^2 = \mathbf{s}^t\,\mathbf{s}$$

- penalize large gradients: finite differencing $\rightarrow \mathbf{v} = \mathbf{H}_v \mathbf{s}$ so

$$\chi^2_{\rm reg} \sim \mathbf{v}^t \mathbf{v} \sim \mathbf{s}^t\,\mathbf{H}_v^t\,\mathbf{H}_v\,\mathbf{s}$$

- penalize large curvature: again finite differencing $\rightarrow$

$$\chi^2_{\rm reg} \sim \mathbf{s}^t\,\mathbf{H}_a^t\,\mathbf{H}_a\,\mathbf{s}$$

all told, use penalty function of the form

$$\chi^2_{\rm reg} \sim \mathbf{s}^t\,\mathbf{H}^t\,\mathbf{H}\,\mathbf{s}$$

with quadratic regularization, full χ^2 is

$$\chi^2 = (\mathbf{L}\,\mathbf{s} - \mathbf{d}^{\rm obs})^t\, \mathbf{S}_d^{-1}\, (\mathbf{L}\,\mathbf{s} - \mathbf{d}^{\rm obs}) + \lambda_s\, \mathbf{s}^t\, \mathbf{H}^t\, \mathbf{H}\, \mathbf{s}$$

where λ_s controls the strength of the regularization:

- low $\lambda \rightarrow$ more emphasis on fit quality
- high $\lambda \rightarrow$ more emphasis on regularization

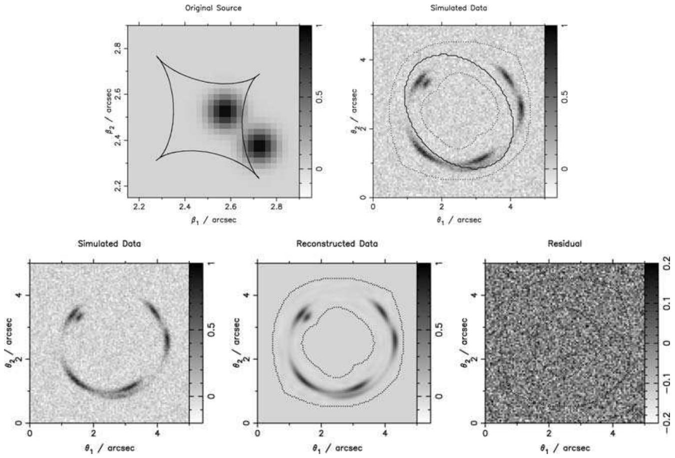
optimal source found analytically — solve $\nabla_{\mathbf{s}}\chi^2 = 0$ or

$$(\mathbf{L}^t\, \mathbf{S}_d^{-1}\, \mathbf{L} + \lambda_s\, \mathbf{H}^t\, \mathbf{H})\, \mathbf{s} = \mathbf{L}^t\, \mathbf{S}_d^{-1}\, \mathbf{d}^{\rm obs}$$

(Warren & Dye 2003; Dye & Warren 2005; Treu & Koopmans 2004; Koopmans 2005; Suyu et al. 2006; Vegetti & Koopmans 2009; coming "soon" to lensmodel)

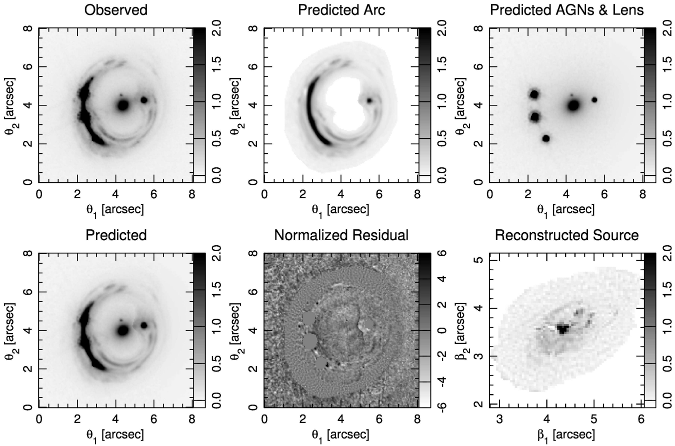
Test case

Suyu et al. 2006MNRAS.371..983S



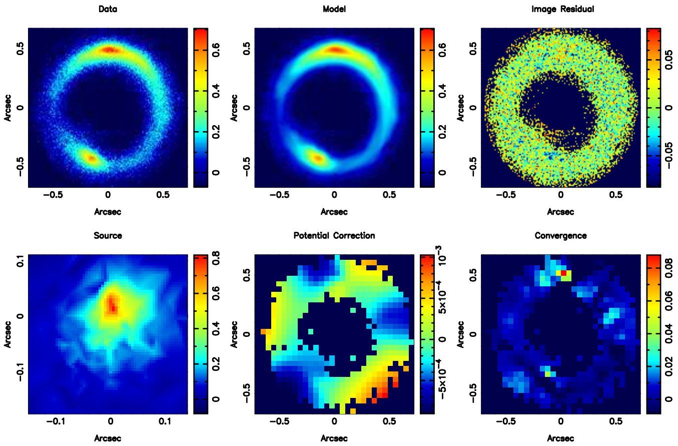
RX J1131−1231

Suyu et al. arXiv:1208.6010



B1938+666

Vegetti et al. 2012Natur.481..341V



Free-form mass models

expand potential or mass in terms of some basis functions

$$\phi(\mathbf{x}) = \sum_\nu a_\nu \, f_\nu(\mathbf{x})$$

parametric vs. non-parametric?

better: over- vs. under-constrained

Constraint equations

theory:

$$\begin{aligned}
\mathbf{u} &= \mathbf{x} - \nabla\phi(\mathbf{x}) \\
\Delta t_{ij} &= t_0 \left[\frac{1}{2} \left(|\mathbf{x}_i - \mathbf{u}|^2 - |\mathbf{x}_j - \mathbf{u}|^2 \right) - \phi(\mathbf{x}_i) + \phi(\mathbf{x}_j) \right] \\
&= t_0 \left[\frac{1}{2} \left(|\mathbf{x}_i|^2 - |\mathbf{x}_j|^2 \right) - (\mathbf{x}_i - \mathbf{x}_j) \cdot \mathbf{u} - \phi(\mathbf{x}_i) + \phi(\mathbf{x}_j) \right]
\end{aligned}$$

constraints from positions and time delays are linear in a_ν, $\mathbf{u}^{\text{mod}}$, and t_0^{-1}:

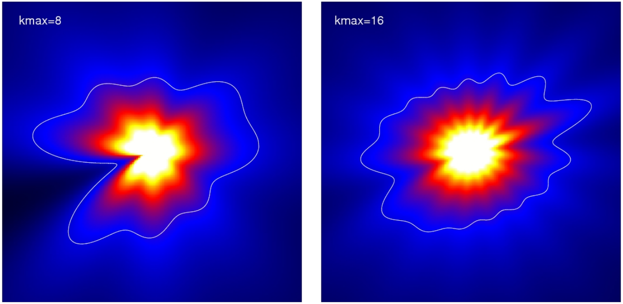
$$\begin{aligned}
\sum_\nu a_\nu \, \nabla f_\nu(\mathbf{x}_i^{\text{obs}}) + \mathbf{u}^{\text{mod}} &= \mathbf{x}_i^{\text{obs}} \\
\Bigg\{ \sum_\nu a_\nu \left[f_\nu(\mathbf{x}_i^{\text{obs}}) - f_\nu(\mathbf{x}_j^{\text{obs}}) \right] + & \\
(\mathbf{x}_i^{\text{obs}} - \mathbf{x}_j^{\text{obs}}) \cdot \mathbf{u}^{\text{mod}} + t_0^{-1} \Delta t_{ij}^{\text{obs}} \Bigg\} &= \frac{1}{2} \left(|\mathbf{x}_i|^2 - |\mathbf{x}_j|^2 \right)
\end{aligned}$$

Multipole models

assume isothermal profile but allow general angular structure

$$\phi_{\rm gal}(r,\theta) = r \sum_{m=0}^{m_{\rm max}} (a_m \cos m\theta + b_m \sin m\theta)$$

apply to a lens with anomalous flux ratios:



(Congdon & CRK 2005; also see Evans & Witt 2003; Yoo et al. 2005, 2006)

Multipole/Taylor models

generalize radial profile — images are often "near" Einstein radius, so do Taylor series expansion in $r - r_0$ (or equivalently $r/r_0 - 1$):

$$\phi(r,\theta) = \sum_{m=0}^{m_{\max}} \sum_{n=0}^{n_{\max}} \left(\frac{r}{r_0} - 1\right)^n (a_{mn} \cos m\theta + b_{mn} \sin m\theta)$$

Trotter et al. 2000ApJ...535..671T apply to MG J0414+0534

Pixelated mass maps

mass pixels — "pixelens" (Saha & Williams 2000, 2004, etc.)

many free parameters — need priors:

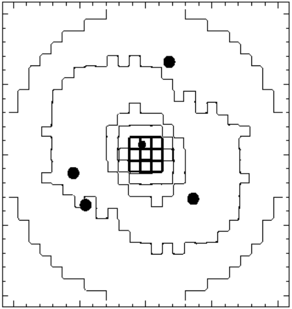
- all pixel densities must be non-negative
- density gradient must point within 45° of lens center
- no pixel value may exceed the average of its neighbors by more than a factor of two (except for central pixel)
- projected density profile must be steeper than $r^{-1/2}$
- if desired, mass map may be required to have inversion symmetry

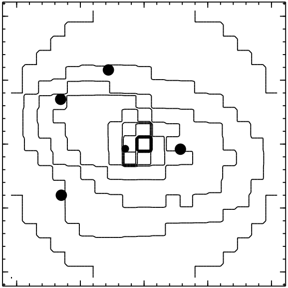
these eliminate models that are grossly unphysical, but are not especially restrictive

- non-negative Σ does not automatically imply non-negative ρ
- no check on number of images
- shapes may still be implausible

examples from Saha & Williams 2004AJ....127.2604S

these show *average* solutions; can also explore range of solutions





Other

other effects I have not gone into...

- pixelated potential corrections

  (Suyu et al. 2009, 2010, 2012; Koopmans 2005; Vegetti et al. 2009, 2010, 2012)

- complicated environments

  (Wong et al. 2011)

- line-of-sight effects (multi-plane lensing)

  (Wong et al. 2011; Suyu et al. 2012)

Bottom line: "precision lensing" is hard work, but we are learning how to do it!