Strong Lens Modeling (I): Principles and Basic Methods

Chuck Keeton

Rutgers, the State University of New Jersey
(I) Principles and Basic Methods

- least-squares fitting
- solving lens equation
- constraints (point data)
- parametric mass models

(II) Statistical Methods

- Bayesian statistics
- Monte Carlo Markov Chains
- nested sampling

(III) Advanced Techniques

- case studies: composite models, astrophysical priors, substructure
- extended sources
- “non-parametric” lens models
Strong lens modeling

goal: use strong lensing data to learn about...  
  ▶ mass model  
  ▶ source  
  ▶ other parameters (e.g., $H_0$)

focus:  
  ▶ galaxy-scale lensing  
  ▶ point data (for now)
Simple examples

“forward” problem:

- fix lens model, solve lens equation to find image positions (and other data)

“inverse” problem:

- fix lens data, (re)interpret lens equation as constraint equation
- solve for model parameters
Point mass

\[ \beta = \theta_1 - \frac{\theta_2^2}{\theta_1} \]

\[ -\beta = \theta_2 - \frac{\theta_2^2}{\theta_2} \]

(double lens; convention: \( \theta_1 > \theta_2 > 0 \))

\[ \theta_1 + \theta_2 = \theta_E^2 \left( \frac{1}{\theta_1} + \frac{1}{\theta_2} \right) \Rightarrow \theta_E = (\theta_1 \theta_2)^{1/2} \]
double lens; again $\theta_1 > \theta_2 > 0$

\[
\beta = \theta_1 - \theta_E \\
-\beta = \theta_2 - \theta_E
\]

then

\[
\theta_E = \frac{\theta_1 + \theta_2}{2} = \frac{\Delta \theta}{2}
\]
Model dependence: Einstein radius

remark: from the same data we can get different answers — depending on what we assume about the models

however ... suppose $\theta_1 = \theta_0 + \delta$ and $\theta_2 = \theta_0 - \delta$, and $\delta$ is small:

\[
\text{ptmass: } \theta_E = \left(\theta_1 \theta_2\right)^{1/2} \approx \theta_0 - \frac{\delta^2}{2\theta_0} + \mathcal{O}(\delta^4)
\]

\[
\text{SIS: } \theta_E = \frac{\theta_1 + \theta_2}{2} = \theta_0
\]

result for Einstein radius is not very sensitive to choice of model

may not be true of other parameters!
**SIS+shear**

lens equation, now in cartesian angular coordinates

\[
\mathbf{u} = \mathbf{x} - \theta_E \hat{\mathbf{x}} - \begin{bmatrix} \gamma x \\ -\gamma y \end{bmatrix}
\]

cross quad: \( u = v = 0 \), with images at \((\pm x_1, 0)\) and \((0, \pm y_2)\)

\[
0 = (1 - \gamma)x_1 - \theta_E \\
0 = (1 + \gamma)y_2 - \theta_E
\]
\[
\begin{align*}
\theta_E + \gamma x_1 &= x_1 \\
\theta_E - \gamma y_2 &= y_2
\end{align*}
\]

then

\[
\begin{bmatrix}
1 & x_1 \\
1 & -y_2
\end{bmatrix}
\begin{bmatrix}
\theta_E \\
\gamma
\end{bmatrix} =
\begin{bmatrix}
x_1 \\
y_2
\end{bmatrix}
\]

solution

\[
\theta_E = \frac{2x_1y_2}{x_1 + y_2}
\quad \text{and} \quad
\gamma = \frac{x_1 - y_2}{x_1 + y_2}
\]
Least-squares fitting

usually we cannot solve the constraint equations exactly
- more constraints than parameters
- noise
- wrong model

general goal: minimize the difference between the model and data

quantify goodness of fit:

$$\chi^2 = \sum \frac{(\text{model} - \text{data})^2}{(\text{uncertainties})^2}$$

idea:
- find best fit (minimum $\chi^2$)
- explore range of allowed models (region where $\chi^2$ is acceptable)
What is “good enough”?

quantify **degrees of freedom**:

\[ \nu = (\# \text{ constraints}) - (\# \text{ free parameters}) \]

if errors are random, have probability distribution for \( \chi^2 \):

\[
p(\chi^2|\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} (\chi^2)^{\nu/2-1} e^{-\chi^2/2}
\]
average:

$$\langle \chi^2 \rangle = \nu$$

peak:

$$\chi_{\text{peak}}^2 = \max(\nu - 2, 0)$$

as a rule of thumb, we expect $\chi^2 \approx \nu$ for a “good” fit; but given statistical scatter, this is not a strict condition!
generalize notion of uncertainties...

if uncertainties are correlated, introduce *covariance*

\[
\text{Cov}(x, y) = \left\langle \left( x - \langle x \rangle \right) \left( y - \langle y \rangle \right) \right\rangle \\
= \left\langle xy - \langle x \rangle y - x \langle y \rangle + \langle x \rangle \langle y \rangle \right\rangle \\
= \langle xy \rangle - \langle x \rangle \langle y \rangle
\]

for an array of data \( d = (d_1, d_2, d_3, \ldots) \), *covariance matrix* is

\[
C = \begin{bmatrix}
\sigma_1^2 & \text{Cov}(d_1, d_2) & \text{Cov}(d_1, d_3) & \cdots \\
\text{Cov}(d_2, d_1) & \sigma_2^2 & \text{Cov}(d_2, d_3) \\
\text{Cov}(d_3, d_1) & \text{Cov}(d_3, d_2) & \sigma_3^2 \\
\vdots & \vdots & \ddots \\
\end{bmatrix}
\]
\[ C = \begin{bmatrix} 0.775 & -0.375 \\ -0.375 & 0.340 \end{bmatrix} \quad \rho_{12} = -0.731 \]

aside: correlation coefficient (dimensionless, \(|\rho| \leq 1|):\]
\[
\rho_{ij} = \frac{\text{Cov}(d_i, d_j)}{\sigma_i \sigma_j}
\]
generalized goodness of fit

\[ \chi^2 = (d_{\text{mod}} - d_{\text{obs}})^t C^{-1} (d_{\text{mod}} - d_{\text{obs}}) \]

if data are independent then

\[ C = \begin{bmatrix} \sigma_1^2 & 0 & \cdots \\ 0 & \sigma_2^2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \]

and \( \chi^2 \) reduces to what you expect

\[ \chi^2 = \sum_i \frac{(d_{\text{mod}}^i - d_{\text{obs}}^i)^2}{\sigma_i^2} \]
example: \( x \) is some independent variable (which we can know); measure \( d^{\text{obs}} \) and postulate a straight line

\[ d^{\text{mod}} = mx + b \]
\[ \chi^2 = \sum_i \left( \frac{m x_i + b - d_i^{obs}}{\sigma_i^2} \right)^2 \]

parabola in both \( m \) and \( b \); find minimum by solving

\[
0 = \frac{\partial \chi^2}{\partial m} = 2 \sum_i \frac{x_i (m x_i + b - d_i^{obs})}{\sigma_i^2} \\
0 = \frac{\partial \chi^2}{\partial b} = 2 \sum_i \frac{(m x_i + b - d_i^{obs})}{\sigma_i^2}
\]

may look complicated, but just a pair of linear equations

\[
\begin{bmatrix}
\sum_i \frac{x_i^2}{\sigma_i^2} & \sum_i \frac{x_i}{\sigma_i^2} \\
\sum_i \frac{x_i}{\sigma_i^2} & \sum_i \frac{1}{\sigma_i^2}
\end{bmatrix}
\begin{bmatrix}
m \\
b
\end{bmatrix} =
\begin{bmatrix}
\sum_i \frac{x_id_i^{obs}}{\sigma_i^2} \\
\sum_i \frac{d_i^{obs}}{\sigma_i^2}
\end{bmatrix}
\]

solve by matrix inversion
Simple Examples
Point mass
SIS
SIS+shear
Least-Squares Fitting
Goodness of fit
Covariance
Linear params
Non-linear params
Linear + non-linear
Errorbars
Solving Lens Eqn
Tiling
Delaunay
lensmodel
Constraints
Positions
Fluxes
Time delays
Parametric Models
Main galaxy
Composite
Environment
Searching
Hands-on
Finding images
Fitting

\[
\begin{bmatrix}
\sum_i \frac{x_i^2}{\sigma_i^2} & \sum_i \frac{x_i}{\sigma_i^2} \\
\sum_i \frac{x_i}{\sigma_i^2} & \sum_i \frac{1}{\sigma_i^2}
\end{bmatrix}
\begin{bmatrix}
m \\
b
\end{bmatrix}
= 
\begin{bmatrix}
\sum_i \frac{x_i d_i^{obs}}{\sigma_i^2} \\
\sum_i \frac{d_i^{obs}}{\sigma_i^2}
\end{bmatrix}
\]

(can generalize to an arbitrary number of linear parameters)
Non-linear parameters

must explicitly search parameter space

use established algorithms to search for minimum of a function in multiple dimensions

challenges:
  ▶ computational effort
  ▶ local minima
  ▶ long, narrow valleys
  ▶ degeneracies
Downhill simplex method ("amoeba")

http://www.cs.usfca.edu/~brooks/papers/amoeba.pdf — also Numerical Recipes
suppose we have parameters $a$ and $b$ such that

\[ d^{\text{mod}} = a f(b) \]

then

\[ \chi^2(a, b) = \sum \frac{(af(b) - d^{\text{obs}})^2}{\sigma^2} \]

optimal value of $a$:

\[ 0 = \frac{\partial \chi^2}{\partial a} = 2 \sum \frac{f(b)(af(b) - d^{\text{obs}})}{\sigma^2} \quad \Rightarrow \quad a_{\text{opt}} = \frac{\sum f(b)d^{\text{obs}}/\sigma^2}{\sum f(b)^2/\sigma^2} \]

then

\[ \chi^2(b) = \chi^2(a_{\text{opt}}(b), b) \]

we can still optimize the linear parameters analytically
Errorbars

“likelihood”
\[ \mathcal{L} \propto e^{-\chi^2/2} \]

1-d Gaussian
\[ \chi^2 = \frac{(x - d)^2}{\sigma^2} \]

\[ \begin{align*}
\pm 1\sigma & : \Delta \chi^2 = 1 \text{ (68\%)} \\
\pm 2\sigma & : \Delta \chi^2 = 4 \text{ (95\%)}
\end{align*} \]

Central region = 68\% of the probability; each tail = 16\%
2-d Gaussian

\[
f = \frac{1}{2\pi \sigma_x \sigma_y} \int <\Delta \chi^2 \exp \left( -\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} \right) \, dx \, dy
\]

\[
= \frac{1}{2\pi} \int <\Delta \chi^2 \exp \left( -\frac{\hat{x}^2 + \hat{y}^2}{2} \right) \, d\hat{x} \, d\hat{y} = \int_0^{\Delta \chi^2} e^{-r^2/2} \, r \, dr
\]

\[
= 1 - e^{-\Delta \chi^2 / 2} \Rightarrow \begin{cases} 
68\% : \Delta \chi^2 = 2.3 \\
95\% : \Delta \chi^2 = 6.2 
\end{cases}
\]
Solving the lens equation

challenges:

- usually non-linear
- often transcendental
- we may not even know how many solutions there are!
  - mathematical theorems bound maximum number of images
  - but we need actual number
- global caustic structure may be informative ... but difficult to find and analyze

solution:

- read lens equation “backwards” — mapping from image position $x$ to unique source position $u(x) = x - \alpha(x)$
- tile image plane
- map each tile back to source plane
- number of tiles that cover source reveals number of images
- tiles themselves give estimates of image positions
Simple Examples
- Point mass
- SIS
- SIS+shear

Least-Squares Fitting
- Goodness of fit
- Covariance
- Linear params
- Non-linear params
- Linear + non-linear
- Errorbars

Solving Lens Eqn
- Tiling
  - Delaunay
  - lensmodel

Constraints
- Positions
- Fluxes
- Time delays

Parametric Models
- Main galaxy
- Composite
- Environment
- Searching

Hands-on
- Finding images
- Fitting
Image plane tiling

- background Cartesian grid — basic coverage
- polar grid centered on each galaxy — resolve key regions
- adaptive subgridding near critical curves
Quadrilaterals vs. triangles

quadrilaterals can be problematic:

triangles are fine:
Delaunay triangulation

start with points in a plane — connect them with triangles

(Google “Delaunay triangulation” — I use http://www.cs.cmu.edu/~quake/triangle.html)
Gridding in gravlens/lensmodel
Magnification and time delay

deflection

\[ \alpha(x) = \nabla \phi(x) = \begin{bmatrix} \phi_x \\ \phi_y \end{bmatrix} \]

magnification

\[
\mu = \det \begin{bmatrix}
1 - \phi_{xx} & -\phi_{xy} \\
-\phi_{xy} & 1 - \phi_{yy}
\end{bmatrix}^{-1} = \left(1 - \phi_{xx}\right)\left(1 - \phi_{yy}\right) - \phi_{xy}^2 \right)^{-1}
\]

special case of circular symmetry, \( \alpha(r) \):

\[
(\text{circular}) \quad \mu = \left[1 - \frac{\alpha(r)}{r}\right]^{-1} \left[1 - \frac{d\alpha}{dr}\right]^{-1}
\]

time delay

\[
t(x; u) = t_0 \left[\frac{1}{2} \left|x - u\right|^2 - \phi(x)\right] \quad t_0 = \frac{1 + z_l}{c} \frac{D_l D_s}{D_{ls}}
\]
Constraints

point sources

data
  ▶ image positions
  ▶ fluxes
  ▶ time delays

source parameters
  ▶ position
  ▶ flux
  ▶ time scale

(extended sources on Thursday)
Position constraints

“exact” position $\chi^2$:

$$\chi^2_{\text{pos}} = \sum_{\text{images}} (x_i^{\text{mod}} - x_i^{\text{obs}})^t S_i^{-1} (x_i^{\text{mod}} - x_i^{\text{obs}})$$

astrometric uncertainties: error ellipse with axes $(\sigma_{1i}, \sigma_{2i})$ and position angle $\theta_{\sigma i}$ (East of North) → covariance matrix

$$S_i = R_i \begin{bmatrix} \sigma_{1i}^2 & 0 \\ 0 & \sigma_{2i}^2 \end{bmatrix} R_i^t \quad R_i = \begin{bmatrix} -\sin \theta_{\sigma i} & -\cos \theta_{\sigma i} \\ \cos \theta_{\sigma i} & -\sin \theta_{\sigma i} \end{bmatrix}$$

if symmetric uncertainties:

$$S_i = \begin{bmatrix} \sigma_i^2 & 0 \\ 0 & \sigma_i^2 \end{bmatrix}$$
note: define source position associated with each observed image

\[ u_{i}^{\text{obs}} = x_{i}^{\text{obs}} - \alpha(x_{i}^{\text{obs}}) \]

also

\[ u^{\text{mod}} = x^{\text{mod}} - \alpha(x^{\text{mod}}) \]

subtract:

\[ \delta u_{i} = \delta x_{i} - \left[ \alpha(x^{\text{mod}}) - \alpha(x_{i}^{\text{obs}}) \right] \approx \mu_{i}^{-1} \cdot \delta x_{i} \]

provided that model is decent, such that \( \delta x_{i} \) and \( \delta u_{i} \) are “small”

then \( \delta x_{i} \approx \mu_{i} \cdot \delta u_{i} \) yields “approximate” position \( \chi^{2} \):

\[ \chi_{\text{pos}}^{2} \approx \sum_{i} (u^{\text{mod}} - u_{i}^{\text{obs}})^{t} \mu_{i}^{t} S_{i}^{-1} \mu_{i} (u^{\text{mod}} - u_{i}^{\text{obs}}) \]
\[
\chi^2_{\text{pos}} \approx \sum_i (\mathbf{u}^\text{mod} - \mathbf{u}_i^\text{obs})^T \mu_i^T \mathbf{S}_i^{-1} \mu_i (\mathbf{u}^\text{mod} - \mathbf{u}_i^\text{obs})
\]

advantages:

- don’t need to solve lens equation
- \(u^\text{mod}\) is a linear parameter, so optimize it analytically

\[
\mathbf{u}^\text{mod} = \mathbf{A}^{-1} \mathbf{b}
\]

where

\[
\mathbf{A} = \sum_i \mu_i^T \mathbf{S}_i^{-1} \mu_i, \quad \mathbf{b} = \sum_i \mu_i^T \mathbf{S}_i^{-1} \mu_i \mathbf{u}_i^\text{obs}
\]

concerns:

- approximation is valid only when residuals are small ... but \(\chi^2_{\text{pos}}\) yields a large value (i.e., bad fit) in either case
- since we do not solve the lens equation, we cannot check that the model predicts correct number of images ... only worry about models yielding \textit{too many} images
**Flux constraints**

\[
\chi^2_{\text{flux}} = \sum_i \frac{(F_i^{\text{obs}} - \mu_i F^{\text{src}})^2}{\sigma_{f,i}^2}
\]

if desired, include parity by letting \( F_i^{\text{obs}} \) and \( \mu_i \) be signed

optimal source flux can be found analytically

\[
F^{\text{src}} = \frac{\sum_i F_i^{\text{obs}} \mu_i / \sigma_{f,i}^2}{\sum_i \mu_i^2 / \sigma_{f,i}^2}
\]

if desired, straightforward to switch to magnitudes

\[
m_i^{\text{mod}} = m^{\text{src}} - 2.5 \log |\mu_i|
\]

note: photometric units are arbitrary — absolute fluxes or magnitudes, or relative values
Time delay constraints

predicted time delay

\[ t^\text{mod}_i = t_0 \tau^\text{mod}_i + T_0 \]

model:

\[ \tau^\text{mod}_i = \frac{1}{2} \left| x^\text{mod}_i - u^\text{mod}_i \right|^2 - \phi \left( x^\text{mod}_i \right) \]

cosmol:

\[ t_0 = \frac{1 + z_l}{c} \frac{D_l D_s}{D_{ls}} = H_0^{-1} \times f(\Omega_M, \Omega_\Lambda; z_l, z_s) \]

note: time zeropoint \( T_0 \) does not affect differential time delays; but let’s make framework general

then

\[ \chi^2_{\text{tdel}} = \sum_i \frac{(t^\text{obs}_i - t_0 \tau^\text{mod}_i - T_0)^2}{\sigma^2_{t,i}} \]
\[
\chi^2_{\text{tdel}} = \sum_i \frac{(t_{\text{obs}}^i - t_0\tau_{\text{mod}}^i - T_0)^2}{\sigma^2_{t,i}}
\]

if we have priors on the cosmological parameters (including \(H_0\))

\[
\rightarrow \text{prior } t_{0,\text{prior}} \pm \sigma_{t_0} \rightarrow \text{additional term}
\]

\[
\chi^2_{t_0} = \frac{(t_0 - t_{0,\text{prior}})^2}{\sigma^2_{t_0}}
\]

optimal values of \(t_0\) and \(T_0\):

\[
\begin{bmatrix}
\sum_i \frac{(\tau_{\text{mod}}^i)^2}{\sigma^2_{t,i}} + \frac{1}{\sigma^2_{t_0}} & \sum_i \frac{\tau_{\text{mod}}^i}{\sigma^2_{t,i}} \\
\sum_i \frac{\tau_{\text{mod}}^i}{\sigma^2_{t,i}} & \sum_i \frac{1}{\sigma^2_{t,i}}
\end{bmatrix}
\begin{bmatrix}
t_0 \\
T_0
\end{bmatrix}
= \begin{bmatrix}
\sum_i \frac{\tau_{\text{mod}}^i t_{\text{obs}}^i}{\sigma^2_{t,i}} + \frac{t_{0,\text{prior}}}{\sigma^2_{t_0}} \\
\sum_i \frac{t_{\text{obs}}^i}{\sigma^2_{t,i}}
\end{bmatrix}
\]
Parametric mass models

postulate: mass distribution can be described by a function with a modest number of parameters

equation: Singular Isothermal Ellipsoid (SIE)

\[ \kappa = \frac{b}{2[(x - x_0)^2 + (y - y_0)^2 / q^2]^{1/2}} \] (+rotation)

pros:
- “easy” to find best fit and assess quality
- build in astrophysical knowledge — assumptions and priors
- “good enough” for many applications

cons:
- can only get out what you put in
- real galaxies may be more complex
# constraints:

<table>
<thead>
<tr>
<th></th>
<th>(x_{gal})</th>
<th>(x_i)</th>
<th>(F_i)</th>
<th>(\Delta t_i)</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>quad</td>
<td>2</td>
<td>(4 \times 2)</td>
<td>4</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>double</td>
<td>2</td>
<td>(2 \times 2)</td>
<td>2</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>

# parameters:

<table>
<thead>
<tr>
<th>(u_{src})</th>
<th>(F_{src})</th>
<th>(x_{gal})</th>
<th>(q_{gal})</th>
<th>(q_{env})</th>
<th>(t_0)</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>(\geq 3)</td>
<td>(\geq 2)</td>
<td>1</td>
<td>(\geq 11)</td>
</tr>
</tbody>
</table>
Main galaxy

softened power law ellipsoid

\[ \kappa = \frac{b^{2-\alpha}}{2(s^2 + x^2 + y^2/q^2)^{1-\alpha/2}} \]

where

\[ M(r) \sim r^\alpha \quad \Rightarrow \quad \alpha \begin{cases} < 1 & \text{steep}er than isothermal} \\ = 1 & \text{isothermal} \\ > 1 & \text{shallower than isothermal} \end{cases} \]

lensmodel has many other model classes: point mass, pseudo-Jaffe, de Vaucouleurs, Hernquist, Sersic, NFW, Nuker, exponential disk, ...
Composite models

can combine multiple components to obtain models that are more complicated but still parametric

for example:
  ▶ stellar component (e.g., Hernquist)
  ▶ dark matter halo (e.g., NFW)

(composite models can be as fancy as you want)
Environmental effects

few lens galaxies are isolated — they have neighbors, and may be embedded in groups or clusters of galaxies

environments can affect the light bending by an amount larger than the measurement uncertainties

if neighboring galaxies are “far” from the lens (compared with Einstein radius), make Taylor series expansion

\[
\phi_{\text{env}} = \phi_0 + \mathbf{a} \cdot \mathbf{x} + \frac{\kappa_c}{2} r^2 + \frac{\gamma}{2} r^2 \cos 2(\theta - \theta_\gamma) \\
+ \frac{\sigma}{4} r^3 \cos(\theta - \theta_\sigma) + \frac{\delta}{6} r^3 \cos 3(\theta - \theta_\delta) + \ldots
\]

structures along the line of sight can also affect the light bending ... more complicated
Searching parameter space

searching parameter space may or may not require a strategic approach...
Hands-on: Finding images

hands-on exercises...

step 1 — pick some mass model, then:

▶ plot grid
▶ plot critical curves and caustics
▶ find images
Hands-on: Fitting

step II — I generated some mock lenses; now you try to fit them

main lens galaxy is a power law ellipsoid

I may have varied:

- mass
- ellipticity/PA
- power law index
- environment: shear/PA, or SIS perturber

all generated with $z_l = 0.3$, $z_s = 2.0$, $\Omega_M = 0.27$, $\Omega_\Lambda = 0.73$, and some fixed value of $H_0$
Sample quads

recall: $z_l = 0.3$, $z_s = 2.0$, $\Omega_M = 0.27$, $\Omega_\Lambda = 0.73$

what are the model parameters? what is $H_0$?
Sample doubles

recall: $z_l = 0.3$, $z_s = 2.0$, $\Omega_M = 0.27$, $\Omega_\Lambda = 0.73$

what are the model parameters? what is $H_0$?