GRAVITATIONAL LENSING AND HIGH REDSHIFT GALAXIES

LECTURE III - WALK THROUGH - BUILD YOUR OWN LENS MODEL

Adi Zitrin

FIRST LIGHT: STARS, GALAXIES AND BLACK HOLES IN THE EPOCH OF REIONIZATION Advanced School, São Paulo

Aug 02

SERIES OF THREE ~90 MIN TALKS:

I. Gravitational lensing

 II. High-redshift galaxies and reionization (through gravitational lenses)

III. "~Hands on" - build your own lens model, luminosity function, and other fun stuff.



MASS MAPPING: STRONG LENSING

GOAL: WE WANT TO CONSTRUCT A LENS MODEL FOR A CLUSTER.

MEANING: WE WANT TO FIT A MODEL TO THE DATA

A DIFFERENT LECTURE: MORE OF A JOINT DISCUSSION

IF THERE'S TIME, WE WILL PLANT SOME SOURCES AND RELENS THEM THROUGH THE LENS

Note: while some of what we will see today is general to lens modeling, what we will do today is a particular (yet true) example to understand the process. There are various lens modeling techniques, parameterizations, pipelines, and methodologies, that one can come up with.

LECTURE III. A LENS MODEL AND LF

We will review:

- 1. Types of models
- 2. Ingredients
- 3. Ray Tracing
- 4. Minimization/optimization
- 5. LF and magnification bias?

HOW TO MODEL A CLUSTER?

What do we want to obtain: mass map, DM, magnification, slope etc.
We need multiple image identifications (and knowledge of the redshifts, distances). These will be our constraints, so a crucial ingredient.



MASS MAPPING: STRONG LENSING

Now generally, there are two typical ways to constrain the model.
The first is perhaps the most elegant and naive. It is called "free-form" or "non-parametric", and basically refers to the fact that one makes no (or very little) assumptions regarding the shape of the underlying mass distribution, and tries to infer it directly from the multiple image constraints.

MASS MAPPING: STRONG LENSING

Now generally, there are two typical ways to constrain the model.

•

- The first is perhaps the most elegant and naive. It is called "free-form" or "non-parametric", and basically refers to the fact that one makes no (or very little) assumptions regarding the shape of the underlying mass distribution, and tries to infer it directly from the multiple image constraints.
 - Using the lens equation, by recalling that the source is the same for each multiple image family, one "solves" for the mass distribution.
- A possible implementation might be setting a grid of NxN pixels on the lens one wishes to model, and then finding the best-fit value of each pixel. If there are say a dozen constraints in an average massive cluster, one can easily see that:

MASS MAPPING: STRONG LENSING

- Now generally, there are two typical ways to constrain the model.
 - The first is perhaps the most elegant and naive. It is called "free-form" or "nonparametric", and basically refers to the fact that one makes no (or very little) assumptions regarding the shape of the underlying mass distribution, and tries to infer it directly from the multiple image constraints.
 - Using the lens equation, by recalling that the source is the same for each multiple image family, one "solves" for the mass distribution.

•

•

A possible implementation might be setting a grid of NxN pixels on the lens one wishes to model, and then finding the best-fit value of each pixel. If there are say a dozen constraints in an average massive cluster, one can easily see that: The problem with "free-form" techniques, despite their novelty, is that often, there are not enough multiple images to learn too much about the exact shape of the mass distribution (in a sense this can be thought of as a very lowresolution result).

FREE-FORM MODELING



FREE-FORM MODELING



MASS MAPPING: STRONG LENSING

 The second method is called "parametric", in which certain parametric, or analytic, forms are assumed for the underlying mass distribution.

MASS MAPPING: STRONG LENSING

The second method is called "parametric", in which certain parametric, or analytic, forms are assumed for the underlying mass distribution.

•

Often a certain analytic form is assumed for the galaxies, following some scaling reactions, and another form for the larger-scale dark matter haloes, where the total deflection field is the superposition of their contributions.

MASS MAPPING: STRONG LENSING

- The second method is called "parametric", in which certain parametric, or analytic, forms are assumed for the underlying mass distribution.
- Often a certain analytic form is assumed for the galaxies, following some scaling reactions, and another form for the larger-scale dark matter haloes, where the total deflection field is the superposition of their contributions.
- Also some model the gas, other structure along the line of sight, etc.

•

Once a general forms (i.e., functions) have been assumed for the different components, an iteration over different values can be made, searching of the values that best-fit the observed features.

MASS MAPPING: STRONG LENSING

- The second method is called "parametric", in which certain parametric, or analytic, forms are assumed for the underlying mass distribution.
- Often a certain analytic form is assumed for the galaxies, following some scaling reactions, and another form for the larger-scale dark matter haloes, where the total deflection field is the superposition of their contributions.
- Also some model the gas, other structure along the line of sight, etc.

•

- Once a general forms (i.e., functions) have been assumed for the different components, an iteration over different values can be made, searching of the values that best-fit the observed features.
 - The advantage is clear: relatively few constraints are needed to obtain the mass model.
- Seems to work very well in both describing observed features in clusters and predict new ones (Treu+2016, Meneghetti+2015). This is what we will do today.

MASS MAPPING: STRONG LENSING

- Before continuing, I'll mention there are also other methods that are not exactly either of the two definitions - such as **Light-Traces-Mass**, in which the dark matter shape is not analytical and does not follow a nice analytic form, but a smoothed version of the clusters galaxies' luminosity distribution (Broadhurst+2005, Zitrin+2009). Or hybrid techniques that include parametric representation of the galaxies but model the dark matter as free form (e.g. **WSLAP+** Diego et al., Sendra et al.).
- OK back to Parametric lens modeling.

•

PARAMETRIC LENS MODELING

• In essence, we wish to find the model (or parameter values describing the model

in the framework of our function choice) that best describes the observations

The idea is clear - let's suppose I have a set of points, and I want to fit a
functional form to it - let's say a linear line, y=ax+b. The goal is to find the best a
and b values that give the minimum "typical" distance between y(x) and y'(x),
where y' is the model (or function's) prediction.

Many of us know this as some sort of Least Squares, for example:



• Now we wish to do the same, but to fit not a linear function, but another function we will construct, where the distance is the distances between the predicted image positions and the observed ones. Will talk about this soon.

QUESTIONS SO FAR?

STEP 1: IDENTIFY MULTIPLE IMAGE CONSTRAINTS

















assume we have photometric or spectroscopic redshifts for them.

WE HAVE THE CONSTRAINTS. WHAT NOW?

WE HAVE THE CONSTRAINTS. WHAT NOW?

We need a sophisticated initial guess of how the mass distribution might be roughly How? as we mentioned, let's first take into account the material we see, the galaxies, then rely on, e.g., simulations to describe the dark matter. Then' we'll adopt some typical parameters and see what happens.

A list of cluster members, their positions and luminosities

How? photometry (run SExtractor or alike), red sequence, photometric-redshifts maybe

PHOTOMETRY



image from http://astroweb.cwru.edu/steven/Virgo/InitialCoreSexCU.html



THE RED SEQUENCE



THE RED SEQUENCE



A list of cluster members, their positions and luminosities

How? photometry (run SExtractor or alike), red sequence, photometric-redshifts maybe

Now, we need to adopt some form for the galaxies - in real life we will want to take ellipticity into account, but here let's go with this: a pseudo isothermal mass distribution:

$$\Sigma(x, y) = \Sigma_0 \frac{r_c r_{cut}}{r_{cut} - r_c} \left(\frac{1}{\sqrt{r_c^2 + \rho^2}} - \frac{1}{\sqrt{r_{cut}^2 + \rho^2}} \right),$$
$$M_{tot} = 2\pi \Sigma_0 r_c r_{cut} = \frac{\pi}{G} \sigma_0^2 r_{cut}^2.$$

Now, we need to adopt some form for the galaxies - in real life we will want to take ellipticity into account, but here let's go with this: a pseudo isothermal mass distribution:

$$\Sigma(x, y) = \Sigma_0 \frac{r_c r_{cut}}{r_{cut} - r_c} \left(\frac{1}{\sqrt{r_c^2 + \rho^2}} - \frac{1}{\sqrt{r_{cut}^2 + \rho^2}} \right),$$

BUT - problem - if we need to account for rc, rcut, and sigma0 of each galaxy we will have also too many free parameters

—> need to adopt some physical or empirical assumptions
 —> can use scaling relations (e.g., from the fundamental plane)

$$\begin{cases} \sigma_0 = \sigma_0^{\star} (\frac{L}{L^{\star}})^{1/4} ,\\ r_{core} = r_{core}^{\star} (\frac{L}{L^{\star}})^{1/2} ,\\ r_{cut} = r_{cut}^{\star} (\frac{L}{L^{\star}})^{\alpha} . \end{cases}$$

Jullo+2007

Mass of each galaxy will be in proportion to its luminosity

Now, we need to adopt some form for the galaxies - in real life we will want to take ellipticity into account, but here let's go with this: a pseudo isothermal mass distribution:

$$\Sigma(x, y) = \Sigma_0 \frac{r_c r_{cut}}{r_{cut} - r_c} \left(\frac{1}{\sqrt{r_c^2 + \rho^2}} - \frac{1}{\sqrt{r_{cut}^2 + \rho^2}} \right),$$

Good - so now with only 3 parameters, we can approximate the contribution form cluster members. Let's choose some reasonable typical values and see how this distribution looks like:

$$\begin{cases} \sigma_0 = \sigma_0^{\star} (\frac{L}{L^{\star}})^{1/4} ,\\ r_{core} = r_{core}^{\star} (\frac{L}{L^{\star}})^{1/2} ,\\ r_{cut} = r_{cut}^{\star} (\frac{L}{L^{\star}})^{\alpha} . \end{cases}$$

Jullo+2007

3 parameters
THE RED SEQUENCE



STEP 3: ACCOUNT FOR DARK MATTER CONTRIBUTION

Choose center, dark matter halo form

Can do also PIEMD! but Let's do NFW



From Wikipedia

2 parameters: either concentration and mass, or, scale radius and central density value

But here we must account for ellipticity and PA, so 2 more Six parameters so far.

Not only the surface mass density of the galaxies and DM are known analytically, also their deflection fields (which we constrain directly) and potential!

This means that alpha_total=alpha_galaxies+alpha_DM simply

THE DARK MATTER MAP





I took some initial values for the dif params:
sigma_0_*~160 km/s
r_c_*~0.3 kpc
DM ell. ~ 0.3
DM PA~45 deg
DM concentration~6
DM mass ~ 3*10^14



MINIMIZATION

So, we have a lens model.

Now how do we know if it is reasonable, or how good it is?

and how do we know of we can better (i.e., if minimized?)

—> Lens multiple images using lens equation, get source position (ray tracing)

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \frac{D_{LS}}{D_{OS}} \boldsymbol{\alpha} \, .$$

—> send again to image plane via lens equation (ray tracing) (why?) —>quantify distance of multiple images from real ones

--> optimize parameters to get minimum distance

MULTIPLE IMAGE PREDICTION



- How to optimize/minimize the parameter values?
- •1. "By Hand"?
- •2. Grid method
- 3. More efficient optimization options

PARAMETRIC LENS MODELING

• In essence, we wish to find the model (or parameter values describing the model

in the framework of our function choice) that best describes the observations

The idea is clear - let's suppose I have a set of points, and I want to fit a
functional form to it - let's say a linear line, y=ax+b. The goal is to find the best a
and b values that give the minimum "typical" distance between y(x) and y'(x),
where y' is the model (or function's) prediction.

Many of us know this as some sort of Least Squares, for example:



• Now we wish to do the same, but to fit not a linear function, but another function we will construct, where the distance is the distances between the predicted image positions and the observed ones.

PARAMETRIC LENS MODELING

Optimization/Minimization

- We need a way to numerically *minimize* the model's deviations from the data.
- In general, one can go over many different a and b parameters, for each combination check what is the sum/average/etc. of deviations, and then choose the "best" a and b as those that yielded the minimum overall deviation. We will want to use such a method.
- There are various ways to draw the different a and b values to probe which are those that result minimum deviation. A simple procedure, for example, can be to go over some a_min to a_max in fixed discrete steps and over b_min to b_max in fixed discrete steps, so that the 2D parameter space is being divided essentially into a grid, where we wish to find the minimum deviation as a function of a and b. Example:

PARAMETRIC LENS MODELING



PARAMETRIC LENS MODELING

Optimization/Minimization

• We can adopt the "grid minimization" method and run over various discrete steps and find the best-fit one. But in fact there are much more efficient algorithms to sample the parameters space and get the best-fit values. You may have heard about some of these method such as downhill simplex, Monte Carlo Markov Chain (MCMC), stimulated annealing, and so forth.

• I am using MCMC with a chi² (ie maximum likelihood) estimator for the goodness of fit (or deviation from the observed multiple image locations).

$$\chi_{SL}^2 = \sum_i \frac{(x'_i - x_i)^2 + (y'_i - y_i)^2}{\sigma_{pos}^2},$$

PARAMETRIC LENS MODELING

Source-versus image plane Minimization

 But we have to be careful. then something funny will happen if I do this naively in the source plane - I will get a very strong lens model that tries to focus everything into a single point. So this procedure is not good because it biases the solution.

 Instead I need to lens the sources back to the image plane and compare to the image position there.

COMMENT: RAY TRACING

- Before we said that we send the images to the source plane, and then we lens back through the lens to the image plane, via the lens equation.
- Is this really that simply?
- theta=beta+alpha*D
- but alpha is known in the image plane, not source plane
- this means one has to run over all the image plane and calculate alpha at each point (time consuming)
- Only pixels that fulfill the lens equation (meaning mapped back to the source)

are counter images



COMMENT: RAY TRACING



COMMMENT: RAY TRACING

- Before we said that we send the images to the source plane, and then we lens back through the lens to the image plane, via the lens equation.
- Is this really that simply?
- theta=beta+alpha*D
- but alpha is known in the image plane, not source plane
- this means one has to run over all the image plane and calculate alpha at each point (time consuming)
- Only pixels that fulfill the lens equation (meaning mapped back to the source)



THE COMBINED MASS MAP! (INITIAL GUESS!)

How to optimize/minimize the parameter

values?

- 1. "By Hand"?
- 2. Grid method
- 3. More efficient optimization options



from: <u>http://bebi103.caltech.edu.s3-website-us-east-1.amazonaws.com/2015/tutorials/</u> <u>t4b_param_est_with_mcmc.html</u>

MINIMIZATION AND THE BAYESIAN APPROACH

The Bayesian approach provides two levels of inference: parameter space exploration, and model comparison. The first level can be achieved using the unnormalised posterior PDF (equal to the product of the likelihood and the prior); the second requires the calculation of the normalisation of the posterior, known as the evidence. All these quantities are related by Bayes Theorem,

$$\Pr(\boldsymbol{\theta}|D, M) = \frac{\Pr(D|\boldsymbol{\theta}, M) \Pr(\boldsymbol{\theta}|M)}{\Pr(D|M)}, \qquad \qquad \mathsf{Jullo+2007} \quad (11)$$

where $\Pr(\boldsymbol{\theta}|D, M)$ is the posterior PDF, $\Pr(D|\boldsymbol{\theta}, M)$ is the likelihood of getting the observed data D given the parameters $\boldsymbol{\theta}$ of the model M, $\Pr(\boldsymbol{\theta}|M)$ is the prior PDF for the parameters, and $\Pr(D|M)$ is the evidence.

The posterior PDF will be the highest for the set of parameters θ which gives the best fit and is consistent with the prior PDF, regardless of the complexity of the model M. Meanwhile, the evidence $\Pr(D|M)$ is the probability of getting the data D given the assumed model M. It measures the complexity of model M, and, when used as in model selection, it acts as <u>Occam's razor</u>: "All things being equal, the simplest solution tends to be the best one." Here, the simplest solution tends to be the model with the smallest number of parameters and with the prior PDF the closest to the posterior PDF. In contrast, the commonly-used reduced χ^2 analysis is only a rough approximation to the evidence analysis, although it does provide an absolute estimator of goodness-of-fit (provided the error estimates on the data are accurate).

MULTIPLE IMAGE PREDICTION: BEST FIT MODEL (UNDER CHOICES MADE)





MULTIPLE IMAGE PREDICTION: BEST FIT MODEL (UNDER CHOICES MADE)



CALCULATE LENS MODEL PROPERTIES

- Now that we have the lens model let's see how we calculate each of the properties:
- Extremely easy!
- Say we start with the deflection field that we fit to the data (alpha_total).
- From this alpha_vector (we have alpha_x and alpha_y separately), kappa, the mass density is simply obtained by kappa=0.5*Div(alpha)
- magnification=abs(1/(1-kappa*2+da_x*da_y-da_x_dy*da_y_dx)); where
- [da_x_dy,da_x_dx]=gradient(alpha_x);
- [da_y_dy,da_y_dx]=gradient(alpha_y);
- Shear:

gamma1=0.5*(da_x_dx-da_y_dy);

gamma2=da_y_dx;

gamma_abs=sqrt(gamma1.^2+gamma2.^2);

• Time Delay:

$$t(\vec{\theta}) = \frac{(1+z_{\rm d})}{c} \frac{D_{\rm d} D_{\rm s}}{D_{\rm ds}} \left[\frac{1}{2} (\vec{\theta} - \vec{\beta})^2 - \psi(\vec{\theta}) \right]$$
$$= t_{\rm geom} + t_{\rm grav} .$$

MAGNIFICATION BIAS

so now that we can do ray tracing, let's try to get the magnification bias together. Let's random sources based on a typical Schechter luminosity function, in the source plane, lens them through the lens, and see what will happen to the number counts given a certain flux limit



LF THROUGH LENSING: HOW TO?

- Suppose we wanted to random a LF, and see the effect of lensing on it. How would we do it?
- Let's make a simple case with a power-law N(S)=AS^-beta
- The idea would be to (i) random sources from the luminosity function and make sure we get what we started with; (ii) random positions in SP and if outside lensed area, do not count source, if in, multiply the flux by the magnification at that position; and (iii) compare to what we started with.
- Easy?
- How would you do (i)?

LENSING OF HIGH REDSHIFT GALAXIES MAGNIFICATION BIAS

To see this let's relens back one FOV we modeled:





Only 44% of original FOV is effectively observed in this case! (not full HST FOV was modeled/lensed, smaller)

LF THROUGH LENSING: HOW TO?

- Suppose we wanted to random a LF, and see the effect of lensing on it. How would we do it?
- Let's make a simple case with a power-law N(S)=AS^-beta
- The idea would be to (i) random sources from the luminosity function and make sure we get what we started with; (ii) random positions in SP and if outside lensed area, do not count source, if in, multiply the flux by the magnification at that position; and (iii) compare to what we started with.
- Easy?
- How would you do (i)?

Data science! HOW TO RANDOM SOURCES FOR A DISTRIBUTION

A trick - the inversion method, generally speaking

• Suppose you have a distribution F(x) you wish to draw from (PDF)

- Get the CDF
- Normalize to range [0,1]
- Uniformly generate random numbers, X, between [0,1]
- Gather CDF values at the positions corresponding to X
- This makes your drawn distribution

Jata science! HOW TO RANDOM SOURCES FOR A DISTRIBUTION

- A trick the inversion method, generally speaking • Suppose you have a distribution F(x) you wish to draw from (PDF)
- Get the CDF
- Normalize to range [0,1]
- Uniformly generate random numbers, X, between [0,1]
- Gather CDF values at the positions corresponding to X
- This makes your drawn distribution (simple!)



Fig from : https://web.mit.edu/urban_or_book/www/book/chapter7/7.1.3.html

see for more details

TAKE HOME BRIEF SUMMARY OF THE THREE TALKS!

 Lecture I: basics of lensing. Main science cases - DM, cosmology, and high redshift galaxies. Based on simple equations (lens equation, alpha, relation to kappa).

TAKE HOME

SUMMARY

- All background is lensed to some extent
- Lensing effects and size of lens depend on mass and on distances, and position of source
- By how much something is lensed? Deflection angle formula for point mass:

$$\hat{lpha} = rac{4GM}{c^2b}$$

- Alpha is constrained with multiple images through lens equation (SL): $\beta =$
- Alpha is then related to the mass distribution through:
- $\vec{\nabla}_{\theta}\vec{\alpha}(\vec{\theta}) = 2\kappa(\vec{\theta})$

$$\theta - \frac{D_{LS}}{D_{OS}} \hat{\alpha}$$

 Shear is constrained in WL regime through ellipticity measurements, also invertible to kappa

Magnification is given by

$$\mu = \frac{1}{\det A} = \frac{1}{\left[(1-\kappa)^2 - \gamma^2\right]}$$

TAKE HOME BRIEF SUMMARY OF THE THREE TALKS!

 Lecture II: lensing of high redshift galaxies. Pushing towards fainter magnitudes. Better internal details. Luminosity function, But tradeoff, magnification bias, completeness simulations.

LENSING OF HIGH REDSHIFT GALAXIES MAGNIFICATION BIAS

To see this let's relens back one FOV we modeled:





Only 44% of original FOV is effectively observed in this case! (not full HST FOV was modeled/lensed, smaller)

LENSING OF HIGH REDSHIFT GALAXIES

MAGNIFICATION BIAS



TAKE HOME

BRIEF SUMMARY OF THE THREE TALKS!

- Lecture III: Building a lens model is something that each of you can do (easily, even , after some guidance/ trials). Just wanted to give a taste of the general idea.
- To do: generate some power law LF, see how slope changes by lensing...

Thank you!