Weak Lensing Lectures

Thomas Kitching (tdk@roe.ac.uk)

May 2011

Notes to accompany Bertinoro school on dark energy, weak lensing lectures. These notes provide a summary of the most important equations and derivations as well as some references to literature for further reading.

All notes and slides available here **http://great10.pbworks.com/Discussion-and-Extra-Material**

- **Part I : Introduction to Lensing**
- **Part II : Basics**
- **Part III : Lensing Measurement**
- **Part IV : Simulations**
- **Part V : Cosmic Shear**
- **Part VI : Predicting Errors**

Part I : Introduction to Lensing

Weak lensing is introduced as a particularly simple, yet powerful probe of cosmology. Here we list some useful reference articles for further reading.

i) Useful References

The papers that first detected the cosmic shear signal were

- Bacon et al., 2000; Mon.Not.Roy.Astron.Soc., 318, 625
- Kaiser et al., 2000; arXiv:astro-ph/0003338
- van Waerbeke et al., 2000; Astron.Astrophys., 358, 30
- Wittman et al., 2000; Nature, 405, 143

References for current dark energy constraints from cosmic shear

- 3D cosmic shear : Kitching et al., (2007); Mon.Not.Roy.Astron.Soc., 376, 771
- 2D cosmic shear : Semboloni et al., (2006); A&A, 452, 51
- 2D cosmic shear : Kilbinger et al. (2009); A&A, 497, 677
- Tomographic cosmic shear : Schrabback et al. (2010); A&A, 516, 63

Useful selected lensing reviews and further information

- Bartelmann & Schneider (2001); Phys.Rept., 340, 472
- Munshi et al. (2008), Phys.Rept., 462, 67
- Heavens (2009); Nuclear Physics B (Proceedings Supplements)
- Massey, Kitching, Richards; (2010), Rep. Prog. Phys., 73, 086901

Part II : Basics

Here we reveiw basic lensing formalism. We derive the lensing distortion amplitude from general relativity and relate this to observable image distortions.

i) Derivation of the Deflection Angle

The lensing equation, that describes the geometric lensing effect is

$$\beta = \theta - \frac{D_{ds}}{D_s}\alpha \tag{1}$$

where β is the observed position of a source and α is the *deflection angle.* This is usually rewritten as

$$\beta = \theta - \hat{\alpha} \tag{2}$$

where $\hat{\alpha}$ is the *reduced deflection angle.*

The reduced deflection angle can be calculated by integrating the line of sight path deflections

$$\hat{\alpha} = \int r_i \mathrm{d}t, \tag{3}$$

this can be calculated from general relativity as follows.

We assume the cosmological principle, and that the time and space parts of the potential are equal such that the metric is given by

$$\mathrm{d}s^2 = -g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = (1+2\Phi)\mathrm{d}t^2 - (1-2\Phi)\delta)\alpha\beta \mathrm{d}x^\alpha \underset{\cdot}{\mathrm{x}}^\beta. \tag{4}$$

We can now write the equation of motion for photons

$$\dot{u}^\mu + \Gamma^\mu_{\nu\lambda} u^\nu u^\lambda = 0 \tag{5}$$

where

$$\Gamma^\mu_{\nu\lambda} = \frac{1}{2} g^{\mu\eta}(g_{\nu\eta,\lambda} + g_{\lambda\eta,\nu} - g_{\nu\lambda,\eta}). \tag{6}$$

We substite Γ into the equation of motion where we can now write down the spatial part (after collecting terms to first order) as

$$\dot{u}^i = -g^{ij}\left[g_{0,j,0} - \frac{1}{2}g_{00,j} + g_{kj,0}u^k + g_{0j,k}u^k - g_{0k,j}u^k + \left(g_{kj,m} - \frac{1}{2}g_{km,j}\right)u^k u^m\right] \tag{7}$$

setting the cross-term parts of the metric to zero $g_{0j} = 0$ we arrive at

$$\dot{u}^i = g^{ij}\left[\frac{1}{2}g_{00,j} - g_{kj,0}u^k - \left(g_{kj,m} - \frac{1}{2}g_{km,j}\right)u^k u^m\right]. \tag{8}$$

We can now substitute the metric in terms of the potential $g_{00} = (1+2\Phi)$ and $g_{ij} = (1-2\Phi)\delta_{ij}$, which leads to

$$\dot{u}^i = 2u^i\dot{\Phi} + \left[(1+u^2)\delta_{ij} - 2u^i u^j\right]\nabla_j\Phi. \tag{9}$$

We can now take the limits of Newtonian gravity and the relativistic limit. In the Newtonian limit we have that $u_i \ll 1$ which leads to $\dot{u}_i = \nabla_i\Phi$. In the relativistic limit $u_i \to r_i$, and we set time derivatives to be zero $\dot{\Phi} = 0$, which leads to $\dot{r}_i = 2(\delta_{ij} - r_i r_j)\nabla_j\Phi$; note that this can be written as a derivative $\dot{r}_i = 2\nabla_\perp\Phi$ perpendicular to the line of sight. Hence the reduced deflection angle can be written like

$$\hat{\alpha} = \int r_i \mathrm{d}t \to 2\int \nabla^\perp \Phi \mathrm{d}t. \tag{10}$$

ii) Weak Lensing Mapping

From the derivation of the deflection angle we can now revisit the lens equation in terms of the potential. By taking the derivative of the lens equation (assuming a linear mapping) we can explore the image distortion from source (θ) to image (β) coordinates

$$\begin{aligned}\left(\frac{\partial\beta}{\partial\theta}\right)_{ij} &= \delta_{ij} - \frac{\partial\alpha}{\partial\theta_i} \\ &= \delta_{ij} - \frac{\partial^2\Psi}{\partial\theta_i\partial\theta_j}\end{aligned} \tag{11}$$

where we have defined a projected Newtonian potential $\Psi = \int \Phi \mathrm{d}z$ called the lensing potential. Note that $\Psi = 0$ implies no potential and no lensing effect.

We define the distortion matrix, that describes the lens coordinate transform as

$$\left(\frac{\partial\beta}{\partial\theta}\right)_{ij} = \delta_{ij} - \frac{\partial^2\Psi}{\partial\theta_i\partial\theta_j} \equiv A_{ij}, \tag{12}$$

which is a matrix that contains elements that are derivatives of the lensing potential Ψ_{11}, Ψ_{12}, Ψ_{21}, Ψ_{22}

$$A_{ij} = \begin{pmatrix} 1-\Psi_{11} & -\Psi_{12} \\ -\Psi_{21} & 1-\Psi_{22} \end{pmatrix}. \tag{13}$$

We identify this as a matrix that can be decomposed into a trace/isotropic part and traceless part. We do this by defining combinations of the potential derivatives that we call *shear* γ and *convergence* κ

$$\begin{aligned}\psi_{11} &= \kappa + \gamma_1 \\ \psi_{22} &= \kappa - \gamma_1 \\ \psi_{21} = \psi_{12} &= \gamma_2,\end{aligned} \tag{14}$$

note that the distortion matrix is symmetric. Using these definitions we can rewrite the distortion matrix like

$$A = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix} = (1-\kappa)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix}, \tag{15}$$

where we have explicitly seperated the trace and traceless parts.

We now want to evaluate what type of distortion this matrix implies. To do this we want to solve the linear set of equations

$$\begin{aligned}(A-\lambda 1)X &= 0 \\ |A-\lambda 1| &= 0\end{aligned} \tag{16}$$

by finding the eigenvalues of these equations. We write the above equation as

$$\begin{vmatrix} 1-\kappa-\gamma_1-\lambda & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1-\lambda \end{vmatrix} = 0, \tag{17}$$

so

$$(1-\kappa-\gamma_1-\lambda)(1-\kappa+\gamma_1-\lambda) - \gamma_2^2 = 0 \tag{18}$$

which results in the quadratic equation

$$\lambda^2 - 2\lambda(1-\kappa) + (1-\kappa)^2 - |\gamma|^2 = 0 \tag{19}$$

which can be solved to find

$$\lambda = (1-\kappa) \pm |\gamma|. \tag{20}$$

These are the semi-major and semi-minor axis of an *ellipse*.

Finally we note that the shear is a *spin-2* quantity (one that is symmetric under rotations of 180 degrees). We can write the two components of the shear field using complex notation where

$$\gamma = \gamma_1 + \mathrm{i}\gamma_2 = |\gamma| \mathrm{e}^{2i\theta}, \tag{21}$$

the factor 2 in the numerator of the exponential notifies you that the shear is a spin-2 field. Equivalently convergence is a scalar field (spin-0); symmetric under any rotation.

The shear and convergence can be easily related to the lensing potential by defining the complex derivative

$$\partial = \partial_x + \mathrm{i}\partial_y \tag{22}$$

where the real and complex derivatives are in the x and y Cartesian coordinates respectively. Using this notation the convergence is written as

$$\kappa = \frac{1}{2}\partial\partial^* \psi \tag{23}$$

and the shear as

$$\gamma = \frac{1}{2}\partial\partial\psi. \tag{24}$$

The projected lensing potential is a scalar field (spin-0) hence by inspection it can be seen that that the complex derivative acts as a raising operator, and its conjugate as a lowering operator: kappa has a raise and lower resulting in a spin-0 field, and shear two raising operators to a spin-2 field.

Part III : Lensing Measurement

We now summarise the two most common methods for measuring the lensing effect.

i) Moments

To measure the lensing effect we can sum over the pixel intensities of galaxies. For example we can find the mean angular position of a distribution of pixel intensities by integrating over angle

$$\bar{\beta} = \frac{\int \mathrm{d}^2\beta q(I(\beta))\beta}{\int \mathrm{d}^2\beta q(I(\beta))}, \tag{25}$$

where $q(I(\beta))$ is some weighted function of the pixel intensities. The mean is the first moment of the pixel distribution, similarly we can calculate the second moment, or *quadrupole* moment of the distribution

$$Q_{ij} = \frac{\int \mathrm{d}\theta_i \mathrm{d}\theta_j q(I(\theta_i,\theta_j))(\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j)}{\int \mathrm{d}\theta_i \mathrm{d}\theta_j q(I(\theta_i,\theta_j))} \tag{26}$$

where we have explicitly written the image angle $\beta = (\theta_i, \theta_j)$.

In a similar way to the distortion matrix we can decompose the quadrupole moment matrix into trace and trace-free parts, where we identify that the matrix describes an ellipticity such that the ellipticity components are

$$(e_1, e_2) = \left(\frac{Q_{11} - Q_{22}}{Q_{11} + Q_{22}}, \frac{2Q_{12}}{Q_{11} + Q_{22}}\right). \tag{27}$$

We now want to ask what is the effect of shear on these moments. To assess with we start by writing the un-lensed (source, S) ellipticity in complex notation as (see Seitz & Schneider, 1995)

$$\chi^S = \frac{(Q^S_{11} - Q^S_{22}) + 2\mathrm{i}Q^S_{12}}{Q^S_{11} + Q^S_{22}}. \tag{28}$$

To calculate the effect of shear we apply the distortion matrix as a rotation of the un-lensed quadrupole moments

$$Q^L = AQA^T. \tag{29}$$

This results in the following equation that relates the observed ellipticity to the unlensed (intrinsic) ellipticity

$$\chi = \frac{\chi^S - 2g + g^2\chi^{S*}}{1 + |g|^2 - 2\mathrm{Re}(g\chi^{S*})} \tag{30}$$

where we introduce the *reduced shear* $g = \gamma/(1-\kappa)$.

There is a similar equivalent transform that uses a different normalisation for the quadrupole moments (see Bonnet & Mellier, 1995)

$$\epsilon = \frac{(Q^S_{11} - Q^S_{22}) + 2\mathrm{i}Q^S_{12}}{Q^S_{11} + Q^S_{22} + 2(Q^S_{11}Q^S_{22} - Q^{S,2}_{12})^{1/2}} \tag{31}$$

which is related to χ by

$$\chi = \frac{2\epsilon}{1 + |\epsilon|^2}. \tag{32}$$

In this case the unlensed ellipticity is related to the observed ellipticity (for $|g| \leq 1$) like

$$\epsilon^S = \frac{\epsilon - g}{1 - g^*\epsilon}, \tag{33}$$

and a similar expression for $|g| > 1$.

We can now take the limit of the ellipticity transformation in the weak lensing limit $|g| \ll 1$, and find that

$$\begin{aligned}\chi &= \chi^S - 2g \\ \epsilon &= \epsilon^S - g\end{aligned} \tag{34}$$

so that the effect of lensing on the ellipticity of an object is a very simple linear addition; and that the two equivalent expressions for ellipticity are simply related by a factor of two.

Here we have introduced the notion of the *intrinsic*, un-lensed ellipticity of a source. In lensing analysis we make the assumption that on average galaxies should be randomly orientated $\langle\epsilon^S\rangle = 0$ so that if we average the observed ellipticities from many galaxies we will recover the shear

$$\begin{aligned}\langle\epsilon\rangle &= \langle\epsilon^S\rangle - \langle g\rangle \\ \langle\epsilon\rangle &\approx \langle g\rangle.\end{aligned} \tag{35}$$

This is a key result that links the observable pixel-level ellipticity to the shear induced by matter along the line of sight enabling us to use weak lensing as a cosmological probe.

ii) Model fitting

An alternative to measuring the moments of the light distributions of galaxies is to create models of the underlying galaxy and fit this model to the data (see Miller et al., 2007 and Kitching et al., 2008). In the most simple case we have a model Galaxy that we convolve with a measured point spread function (PSF) and calculate the likelihood over the data for each galaxy where the ellipticity of the galaxy (e_1,e_2) and other parameters $\boldsymbol{\theta}$ (for example the size, brightness, position) are free parameters

$$\mathcal{L} = \ln L(e_1, e_2, \boldsymbol{\theta}) = \sum_{\rm d,data} ({\rm Galaxy} * {\rm PSF} - {\rm Data})C^{-1}({\rm Galaxy} * {\rm PSF} - {\rm Data})^T. \tag{36}$$

To arrive at a probailistic measure of the ellipticity we then marginalise over the other parameters

$$L(e_1, e_2) = \int {\rm d}\theta_\alpha L(e_1, e_2, \theta_\alpha). \tag{37}$$

To work in a consistent Bayesian formalism we can write a posterior $P_g(e)$ for the ellipticity of each galaxy by using a prior on the intrinsic ellipticity distribution $p(e)$ such that

$$P_g(e|y) = \frac{p(e)L_g(y|e)}{\int p(e')L_g(y|e'){\rm d}e'} \tag{38}$$

where y is the data.

We know from the moments discussion that the mean ellipticity averaged over galaxies is an approximation of the shear. In the probabilistic formalism of model fitting this can be written the expectation value of the ellipticity

$$\langle e\rangle = \frac{1}{N}\sum_i^N \int eP_g(e|y_i){\rm d}e = \frac{1}{N}\int e\sum_i^N P_g(e|y_i){\rm d}e, \tag{39}$$

over N galaxies. This is the expectation value of the observed ellipticity.

There is an additional effect that must be account for in model fitting (and moments, though in that case it is harder to account for), because a low signal-to-noise the likelihood from a galaxy will be very broad (very lose contraints on ellipticity), and posterior will be dominated by the prior, hence the expectation value of the ellipticity will be biased. In a Bayesian formalism however this bias can be exactly accounted for. To account for this 'noise bias' we write the expectation value of the ellipticity as a Taylor expansion such that

$$\langle e_1 \rangle_i \approx e^S_{1,i} + g_1 \partial \langle e_1 \rangle_i / \partial g_1 + g_2 \partial \langle e_1 \rangle_i / \partial g_2 + \dots \tag{40}$$

from this it can be seen that the average of the expectation value over an ensemble of galaxies is

$$\sum_i^N \langle e_1 \rangle_i \approx g_1 \sum_i^N \partial \langle e_1 \rangle_i / \partial g_1 \tag{41}$$

so that

$$\hat{g}_\mu = \frac{\sum_i^N w_i \langle e_\mu \rangle_i}{\sum_i^N w_i \partial \langle e_\mu \rangle_i / \partial g_\mu} \tag{42}$$

where we generalise to both ellipticity components $\mu = (1,2)$ and we include a generic weight w_i that could be optimised. The dominator is called the *sensitivity* and can be calculated exactly in the model fitting Bayesian approach (seee Miller et al., 2007 and Kitchign et al., 2008).

Part IV : Simulations

Because we can never observe the unlensed ellipticity of objects algorithms that attempt to measure the shear must be tested against simulations. In these simulations a set of simulated galaxies is sheared by a known amount and this 'true/simulated shear' is compared to the measured shear provided by the algorithms.

There are four publically available lensing simulations from two related programmes STEP (the Shear TEsting Programme) and GREAT (the GRavitational lEnsing Accuracy Testing) more information can be found here `http://www.greatchallenges.info` we summarise the main features of these simulations here

- STEP 1 : Heymans et al., 2006. Simple galaxy models, randomly distributed objects, unknown PSF and constant shear.
- STEP 2 : Massey et al., 2007. Complex galaxy models, randomly distributed objects, unknown PSF and constant shear.
- GREAT08 : Bridle et al., 2009,2010. Simple galaxy models, known galaxy positions, known PSF, constant shear.
- GREAT10 : Kitching et al., 2011. On going challenge (c. May 2011), known galaxy positions, known variable PSF, variable shear.

Also available, but not published are the STEP 3 (like STEP 2 but with different PSFs) and STEP 4 (like GREAT08) simulated data sets.

Part V : Cosmic Shear

Because the mean of lensing shear effect is zero when averaged over a sufficiently large number of galaxies (under the assumption of the cosmological principle) it is the variance of the shear field that contains cosmological information; which is encapsulated in the correlation function or power spectrum of the shear (the correlation function is the Fourier transform of the shear). Here we will derive the 3D weak lensing power spectrum, which is analogous to the 2D $C(\ell)$ CMB power spectrum; weakly lensed galaxies have redshift and shear information hence we need to work with a 3D power spectrum.

The following outlines the derivations in Heavens et al. (2006) and Kitching et al., (2010). To start we show a general spherical harmonic transform, the 3D spin-weight equivalent of a Fourier transform

$$ {}_sf_{\ell m}(k) \equiv \sqrt{\frac{2}{\pi}} \int \mathrm{d}^3 r\, {}_sf(r) k j_\ell(kr)\, {}_sY_\ell^{m*}(\hat{n}), \tag{43} $$

where Y are are spherical spin-weight polynomials, and j are Bessel functions. k are radial wavevectors and ℓ and m are angular wavevectors. In the flat sky approximation for a scalar field $(s = 0)$ this reduces to

$$ f(k, \ell) \equiv \sqrt{\frac{2}{\pi}} \int \mathrm{d}^3 r f(r) k j_\ell(kr) \exp(\mathrm{i}\ell.\theta). \tag{44} $$

The covariance of these flat sky coefficients is related to the power spectrum of f

$$ \langle f(k,\ell) f^*(k',\ell')\rangle = (2\pi)^2 P_f(k) \delta^D(k - k') \delta^D(\ell - \ell') \tag{45} $$

where δ^D is the Dirac delta function.

For the 3D shear field we can now write an estimator for the spherical harmonic coefficients for data,

$$ \hat{\gamma}(k, \ell) = \sqrt{\frac{2}{\pi}} \sum_g \gamma_g(r) k j_\ell(kr(z_g)) \exp(\mathrm{i}\ell.\theta) \tag{46} $$

where the sum is over all galaxies. We would like to measure this from data and compare the covariance of the coefficients with an expected covariance from theory.

To generate a theoretical covariance we start by writing down the set of coefficients

$$ \hat{\gamma}(k, \ell) \equiv \sqrt{\frac{2}{\pi}} \int \mathrm{d}^3 r n(r) \gamma(r) k j_\ell(kr) \exp(\mathrm{i}\ell.\theta). \tag{47} $$

where the integrated shear at each comoving distance is weighted by the number density at that distance. From earlier (Part I) we know that the shear is related to the lensing potential via $\gamma(r) = (1/2)\partial\partial\phi(r)$ and that the lensing potential is related to the Newtonian potential like

$$ \phi(r) = \int_0^r \mathrm{d}r' F_K(r, r') \Phi(r') \tag{48} $$

where the $F_K(r, r') = D_{rr'}/D_r$ convert from reduced deflection angle and the integral is along the line of sight.

We can related the Newtonian potential to the local matter overdensity $\delta \equiv [\rho(r) - \bar{\rho}]/\bar{\rho}$ by Poissons equation

$$ \nabla^2 \Phi = \frac{3\Omega_m H_0^2}{2a(t)} \delta \tag{49} $$

or in Fourier space

$$\Phi(k,\ell) = -\frac{3\Omega_m H_0^2}{2a(t)k^2}\delta(k,\ell). \tag{50}$$

We now substitute Poissons equation in Fourier space and transform back to real space, then substitute the second derivative of the lensing potential into the equation for the shear coefficients which leads to

$$\begin{aligned}\hat{\gamma}(k,\ell) = -\frac{3\Omega_m H_0^2}{2\pi c^2}\int \mathrm{d}z \\ \int \mathrm{d}^2\theta k j_\ell(k,r)\bar{n}(z)\exp(-\mathrm{i}\ell.\theta) \\ \int_0^r \mathrm{d}r' a^{-1}(r')F_K(r,r') \\ \int \mathrm{d}k'\frac{\mathrm{d}^2\ell'}{2\pi}k' j_{\ell'}(k'r')\delta(k',\ell';r')\frac{X_\ell}{2}\exp(\mathrm{i}\ell'.\theta)\end{aligned} \tag{51}$$

where we have replaced a 3D integral over space with an integral over angle and redshift, and the derivatives of $\phi(r)$ over θ generate act on the numerator of the exponentials to generate the function X_ℓ. By integrating over θ a delta function in ℓ is generated which simplifies the expression further to

$$\begin{aligned}\hat{\gamma}(k,\ell) = -\frac{3\Omega_m H_0^2 X_\ell}{4\pi c^2}\int \mathrm{d}z j_\ell(k,r)\bar{n}(z) \\ \int_0^r \mathrm{d}r' a^{-1}(r')F_K(r,r') \\ \int \mathrm{d}k' k' j_{\ell'}(k'r')\delta(k',\ell';r').\end{aligned} \tag{52}$$

This expression is a theoretical expression for the shear transform coefficients.

We now take the covariance of these estimators $\langle\gamma(k,\ell)\gamma^*(k',\ell')\rangle$ to generate the power spectrum. In doing this we make use of the expression

$$\langle\delta(k,\ell;r)\delta^*(k',\ell';r')\rangle \approx (2\pi)^2\sqrt{P_\delta(k;r)P_\delta(k';r')}\delta^D(k-k')\delta^D(\ell-\ell') \tag{53}$$

which is a geometric approximation of matter overdensity correlations, where P_δ is the matter power spectrum.

The final expression for the 3D shear power spectrum is therefore

$$\begin{aligned}C_\ell(k_1,k_2) = \mathcal{A}^2\int \mathrm{d}r_g r_g^2 n(r_g) j_\ell(k_1 r_g)\int \mathrm{d}r_h r_h^2 n(r_h) j_\ell(k_2 r_h) \\ \int \mathrm{d}\tilde{r}'\int \mathrm{d}\tilde{r}''\frac{F_K(r_g,\tilde{r}')}{a(\tilde{r}')}\frac{F_K(r_h,\tilde{r}'')}{a(\tilde{r}'')}\int\frac{\mathrm{d}k'}{k'^2}j_\ell(k'\tilde{r}')j_\ell(k'\tilde{r}'')P_\delta^{1/2}(k';\tilde{r}')P_\delta^{1/2}(k';\tilde{r}''),\end{aligned} \tag{54}$$

this clearly contains geometric terms (the F_K) and large scale structure terms (the P_δ), and shows how the observable lensing distortion is related to the underlying cosmological functions.

The former expression is a continuous estimator of the power spectrum, however a more commonly used quantity is the 'tomographic' power spectrum which is a series of 2D power spectrum binned in redshift. The tomographic power spectrum is related to the 3D cosmic shear power spectrum via three transformations

1. The Limber approximation.

2. A Fourier-space to real space transform.
3. A discretisation of the shear field by binning in redshift or comoving distance.

The Limber approximation can be thought of as a projection in wavespace from a 3D wavevector (k_x, k_y, k_z) to a 2D wavevector (k_x, k_y). LoVerde & Afshordi (2008) provided a very useful formula to implement the Limber approximation that simply replaces Bessel functions by delta functions in the limit of small scales

$$\lim_{\ell\to\infty} j_\ell(kr) \to \sqrt{\frac{\pi}{2(\ell+\frac{1}{2})}}\delta^D\left(kr - \left[\ell + \frac{1}{2}\right]\right). \tag{55}$$

By applying this limit to the 3D cosmic shear field we arrive at the expression

$$C_\ell^{\rm Limber}(k_1, k_2) = \frac{9\Omega_m^2 H^4}{4c^2}\int {\rm d}r \frac{P(\ell/r; r)}{a(r)^2}\frac{W(r_1, r)W(r_2, r)}{r^2} \tag{56}$$

by converting to real space performing a discretisation in redshift space we arrive at the most commonly used form for the cosmic shear power spectrum

$$C(\ell)_{ij} = \int {\rm d}r W_{ij}^{GG} P(\ell/r; r) \tag{57}$$

where the indices ij now refer to redshift bins and

$$W_{ij}^{GG} = \frac{q_i(r)q_j(r)}{r^2} \tag{58}$$

with

$$q_i(r) = \frac{3H^2\Omega_m r}{2a(r)}\int {\rm d}r' n(z)\frac{(r-r')}{r'}. \tag{59}$$

The correlations that are taken to generate the tomographic power are 2D correlations that are done either in the same redshift bin "auto-correlations" where $i = j$ or between bins "cross-correlations" where $i \neq j$.

Part VI : Predicting Errors[1]

Fisher matrices are an extremely useful tool in experimental design, they allow for the potential accuracy of an experiment to be assessed before it is even built!

i) *Approximating the Likelihood Surface*

Let us assume we have a posterior probability distribution, which is single-peaked. Two common estimators (indicated by a hat: $\hat{\theta}$) of the parameters are the peak (most probable) values, or the mean,

$$\hat{\theta} = \int d\theta\, \theta\, p(\theta|\boldsymbol{x}). \tag{60}$$

An estimator is *unbiased* if its expectation value is the true value θ_0:

$$\langle \hat{\theta} \rangle = \theta_0. \tag{61}$$

Let us assume for now that the prior is flat, so the posterior is proportional to the likelihood. This can be relaxed. Close to the peak, a Taylor expansion of the log likelihood implies that locally it is a mutivariate gaussian *in parameter space*:

$$\ln L(\boldsymbol{x};\theta) = \ln L(\boldsymbol{x};\theta_0) + \frac{1}{2}(\theta_\alpha - \theta_{0\alpha})\frac{\partial^2 \ln L}{\partial\theta_\alpha \partial\theta_\beta}(\theta_\beta - \theta_{0\beta}) + \ldots \tag{62}$$

where, at the peak the gradient is zero $\ln L_\alpha = 0$. This can be also be expressed as

$$L(\boldsymbol{x};\theta) = L(\boldsymbol{x};\theta_0)\exp\left[-\frac{1}{2}(\theta_\alpha - \theta_{0\alpha})H_{\alpha\beta}(\theta_\beta - \theta_{0\beta})\right]. \tag{63}$$

The Hessian matrix $H_{\alpha\beta} \equiv -\frac{\partial^2 \ln L}{\partial\theta_\alpha\partial\theta_\beta}$ controls whether the estimates of θ_α and θ_β are correlated or not. If it is diagonal, the estimates are uncorrelated. Note that this is a statement about *estimates* of the quantities, not the quantities themselves, which may be entirely independent, but if they have a similar effect on the data, their estimates may be correlated. Note that in cases of practical interest, the likelihood may not be well described by a multivariate gaussian at levels which s et the interesting credibility levels (e.g. 68%). We turn later to how to proceed in such cases.

ii) *Conditional Errors*

If we fix all the parameters except one, then the error is given by the curvature along a line through the likelihood (posterior, if prior is not flat):

$$\sigma_{\text{conditional},\alpha} = \frac{1}{\sqrt{H_{\alpha\alpha}}}. \tag{64}$$

This is called the *conditional error*, and is the minimum error bar attainable on θ_α if all the other parameters are known. *It is rarely relevant and should almost never be quoted.*

iii) *Marginalising and Marginalised Errors*

The marginal distribution of for a parameter θ_1 is obtained by integrating the likelihood over the other parameters:

$$p(\theta_1) = \int d\theta_2 \ldots d\theta_N p(\theta) \tag{65}$$

a process which is called *marginalisation*. Often one sees marginal distributions of all parameters in pairs, as a way to present some complex results. In this case one variable is left out of the integration.

[1]Thanks to Alan Heavens for providing some latex for this Section

If you plot such error ellipses, you *must* say what contours you plot. If you say you plot 1σ and 2σ contours, this could be for the joint distribution (i.e. 68% of the probability lies within the inner contour), or the 68% probability of a single parameter that lies within the bounds projected onto a parameter axis. The latter is a 1σ, single-parameter error contour (and corresponds to $\Delta\chi^2 = 1$), whereas the former is a 1σ contour for the joint distribution, and corresponds to $\Delta\chi^2 = 2.3$. **Read Numerical Recipies Section 15.6 (really, read this now!)**.

A multivariate gaussian likelihood is a common assumption, so it is useful to compute marginal errors for this rather general situation. The simple result is that the marginal error on parameter θ_α is

$$\sigma_\alpha = \sqrt{(H^{-1})_{\alpha\alpha}}. \tag{66}$$

Note that we invert the Hessian matrix, and then take the square root of the diagonal components.

iv) *The Fisher (Information) Matrix*

How accurately can we estimate model parameters from a given data set? This question was basically answered 75 years ago Fisher (1935), and we will now summarize the results, which are both simple and useful.

If we are fitting a model to some data we want the estimate of parameters to be unbiased, i.e.,

$$\langle\theta\rangle = \theta_0, \tag{67}$$

and give as small error bars as possible, i.e. , minimize the standard deviations

$$\Delta\theta_\alpha \equiv \left(\langle\theta_\alpha^2\rangle - \langle\theta_\alpha\rangle^2\right)^{1/2}. \tag{68}$$

In statistics jargon, we want the BUE θ_α, which stands for the "Best Unbiased Estimator".

A key quantity is the the *Fisher Matrix* (or Information matrix) that is defined as the expectation of the Hessian matrix

$$F_{\alpha\beta} \equiv \langle H_{\alpha\beta}\rangle = \left\langle -\frac{\partial^2 \ln L}{\partial\theta_\alpha \partial\theta_\beta}\right\rangle. \tag{69}$$

Another key quantity is the *maximum likelihood estimator*, or *ML-estimator* for brevity, defined as the parameter vector θ that maximizes the likelihood function $L(x;\theta)$.

A number of powerful theorems have been proven

1. For any unbiased estimator, $\Delta\theta_\alpha \geq 1/\sqrt{F_{\alpha\alpha}}$ (the *Cramér-Rao* inequality).
2. If an unbiased estimator attaining ("saturating") the Cramér-Rao bound exists, it is the ML estimator (or a function thereof).
3. The ML-estimator is asymptotically BUE.

The first of these theorems, known as the Cramér-Rao inequality, thus places a firm lower limit on the error bars that one can attain, regardless of which method one is using to estimate the parameters from the data.

The normal case is that the other parameters are estimated from the data as well, in which case the minimum standard deviation rises to

$$\Delta\theta_\alpha \geq (F^{-1})^{1/2}_{\alpha\alpha}. \tag{70}$$

This is called the *marginal error*, and is normally the relevant error to quote. It is always at least as large as the expected conditional error $1/F_{\alpha\alpha}$.

v) *The Gaussian Case*

Let us now explicitly compute the Fisher information matrix for the case when the probability distribution is Gaussian, *i.e.*, where (dropping an irrelevant additive constant $N\ln[2\pi]$)

$$2\mathcal{L} = \ln\det C + (\boldsymbol{x}-\mu)C^{-1}(\boldsymbol{x}-\mu)^T, \tag{71}$$

where in general both the mean vector μ and the covariance matrix

$$C = \langle(\boldsymbol{x}-\mu)(\boldsymbol{x}-\mu)^T\rangle \tag{72}$$

depend on the model parameters $\boldsymbol{\theta}$. Although vastly simpler than the most general situation, the Gaussian case is nonetheless general enough to be applicable to a wide variety of problems in cosmology. Defining the data matrix

$$D \equiv (\boldsymbol{x}-\mu)(\boldsymbol{x}-\mu)^T \tag{73}$$

and using the matrix identity (see exercises) $\ln\det C = \mathrm{Tr}\ln C$, where Tr indicates trace, we can re-write (71) as

$$2\mathcal{L} = \mathrm{Tr}\left[\ln C + C^{-1}D\right]. \tag{74}$$

We will use the standard comma notation for derivatives, where for instance

$$C_{,\alpha} \equiv \frac{\partial}{\partial\theta_\alpha}C. \tag{75}$$

Since C is a symmetric matrix for all values of the parameters, it is easy to see that all the derivatives $C_{,\alpha}$, $C_{,\alpha\beta}$, will also be symmetric matrices. Using the matrix identities $(C^{-1})_{,\alpha} = -C^{-1}C_{,\alpha}C^{-1}$ and $(\ln C)_{,\alpha} = C^{-1}C_{,\alpha}$ (see exercises), we find

$$2\mathcal{L}_{,\alpha} = \mathrm{Tr}\left[C^{-1}C_{,\alpha} - C^{-1}C_{,\alpha}C^{-1}D + C^{-1}D_{,\alpha}\right]. \tag{76}$$

When evaluating C and μ at the true parameter values, we have $\langle\boldsymbol{x}\rangle = \mu$ and $\langle\boldsymbol{x}\boldsymbol{x}^T\rangle = C + \mu\mu^T$, which gives

$$\begin{cases} \langle D\rangle & = C, \\ \langle D_{,\alpha}\rangle & = 0, \\ \langle D_{,\alpha\beta}\rangle & = \mu_{,\alpha}\mu_{,\beta}^T + \mu_{,\beta}\mu_{,\alpha}^T. \end{cases} \tag{77}$$

Using this and equation (76), we obtain $\langle\mathcal{L}_{,\alpha}\rangle = 0$. In other words, the ML-estimate is correct on average in the sense that the average slope of the likelihood function is zero at the point corresponding to the true parameter values. Applying the chain rule to equation (76), we obtain

$$\begin{aligned} 2\mathcal{L}_{,\alpha\beta} = \mathrm{Tr}[\; & - \; C^{-1}C_{,\alpha}C^{-1}C_{,\beta} + C^{-1}C_{,\alpha\beta} \\ & + \; C^{-1}(C_{,\alpha}C^{-1}C_{,\beta} + C_{,\beta}C^{-1}C_{,\alpha})C^{-1}D \\ & - \; C^{-1}(C_{,\alpha}C^{-1}D_{,\beta} + C_{,\beta}C^{-1}D_{,\alpha}) \\ & - \; C^{-1}C_{,\alpha\beta}C^{-1}D + C^{-1}D_{,\alpha\beta}]. \end{aligned} \tag{78}$$

Substituting this and equation (77) into equation (**??**) and using the trace identity $\mathrm{Tr}[AB] = \mathrm{Tr}[BA]$, many terms drop out and the Fisher information matrix reduces to simply

$$F_{\alpha\beta} = \langle\mathcal{L}_{,\alpha\beta}\rangle = \frac{1}{2}\mathrm{Tr}[C^{-1}C_{,\alpha}C^{-1}C_{,\beta} + C^{-1}M_{\alpha\beta}], \tag{79}$$

where we have defined the matrix $M_{\alpha\beta} \equiv \langle D_{,\alpha\beta}\rangle = \mu_{,\alpha}\mu_{,\beta}^T + \mu_{,\beta}\mu_{,\alpha}^T$.

This result is extremely powerful. If the data have a (multivariate) gaussian distribution (and the errors can be correlated; C need not be diagonal), and you know how the means μ and the covariance matrix C depend on the parameters, you can calculate the Fisher Matrix *before you do the experiment*. The Fisher Matrix gives you the expected errors, so you know how well you can expect to do if you do a particular experiment, and you can then design an experiment to give you, for example, the best (marginal) error on the parameter you are most interested in.

vi) *Fisher Matrix Operations*

Adding extra parameters. If we start with a parameter set θ and we wish to extend this to include some extra parameter ϕ then the new Fisher matrix is extended by adding a block diagonal element containing the ϕ parameters and two blocks containing the cross-correlation between θ and ϕ.

$$\begin{matrix} F^{\theta\theta} & F^{\theta\phi} \\ F^{\phi\theta} & F^{\phi\phi} \end{matrix} \tag{80}$$

For single parameters this is equivalent to simply adding an extra row and column to the original matrix.

Marginalising over parameters in the set. If we have a parameter set $\theta + \phi$ and we wish to marginalise over the parameters ϕ we can construct the Schur compliment in θ

$$F_S^{\theta\theta} = F^{\theta\theta} - F^{\theta\phi}(F^{\phi\phi})^{-1}F^{\phi\theta} \tag{81}$$

The errors calculated from the new matrix F_S will now include a marginalisation over all parameter in ϕ. This is equivalent to the following operations : i) invert the full Fisher matrix, ii) remove the ϕ rows and columns from the inverse matrix to make a θ-only inverse, iii) re-invert the smaller θ-only matrix.

Changing parameter set (rotation). Given a set of parameters θ we may want to change parameter set. To do this we can rotate the Fisher matrix using a Jacobian transformation

$$F_R^{\xi\xi} = J^T F^{\theta\theta} J \tag{82}$$

where J is a rotation vector that contains derivative of the new parameters with respect to the old one $J = \partial\xi/\partial\theta$. Note that the ξ parameters maust be a linear combination of the θ parameters, if this is not the case then the transformation will not be an affine (loss-less) transformation. There is a special case when the new F^R is diagonal, in this circumstance the transformation is call an *eigenvalue decomposition*; and the rows of the rotation matrix are refered to as eigenvectors. This type of rotation is equivalent to a rotation in parameter space – then eigenvalue decomposition is the case where the rotation produces un-correlated error ellipises.

Combining independant experiments. If n experiments are independent then the combined constraints on parameters can be easily calculated simply by adding the Fisher matrices, though care must be taken to add the correct rows and columns. If the experiments are not independent (i.e. there is some correlation between the signals) then a single Fisher matrix mucst be constructed from the beginning where the covariance used to constraint the parameters contains the relevant correlation.

vii) *iCosmo Fisher Matrices – Lensing Example*

iCosmo `http://www.icosmo.org` is an open source code in IDL that can be used to generate cosmological Fisher matrices.

As example of how Fisher matrices are calculated in iCosmo we will use the example of the weak lensing tomography Fisher matrix. To calculate Fisher matrices we need to know the derivative of the mean (and/or covariance) with respect to cosmological parameters; in the case of weak lensing tomography the power spectra can be thought of as the mean. Since the derivatives of these complex functions is not analytic (except in rare cases) we must evaluate the derivatives numerically. In iCosmo there are two functions that can calculate the derivatives of the lensing power spectra with an increasing order of complexity.

`mk_dcldp_2p(f,s,'param')`. This calculates the derivative of the lensing power spectra using a two step, linear, approximation. The arguments are the fiducial and survey structures (using a lensing survey) as well as a string indicating which parameter should be varied in the derivative; this can be for example 'omega_m', 'w0', 'ns' etc. The output structure contains the lensing power spectrum derivative as a function of ℓ and redshift.

`mk_dcldp_4p(f,s,'param')`. This is exactly the same as `mk_dcldp_2p` except that the derivative is calculated using a four step approximation, a quadratic approximation to the likelihood. Note that twice the number of points in parameter space are calculated so the code will take twice the time to run.

Finally to cacluate the Fisher matrix the noise (error) on the lensing power spectra must be evaluated and the Fisher matrix elements themselves created. In iCosmo all these calculations, as well as the parameter derivatives are performed within a single function.

`mk_fisher_lens(f,s)`. This calculates the lensing Fisher matrix using the fiducial and survey structures. The output Fisher structure contains the Fisher matrix, its inverse (`fish.cov`), as well as the marginal and conditional errors. The code also performs a check to make sure that the Fisher matrix is positive definite (done using `check_matrix`) the status variable in the Fisher structure reflects the validity of this check.

To add two fishers use the `comb_fisher` command, this will automatically add the correct rows and columns corresponding to each parameter.

viii) *Plotting Fisher Matrices*

In order to plot a 2D ellipse in iCosmo we must first marginalise over all parameters except thos that are being plotted. This is done using the `margin_fisher` routine which takes the fisher and a string of yes (1) or no (0) operators and will create a new Fisher for which the parameters that are not chosen (0) are marginalised over for example

```
> margin_fisher,fishin,fishout,[0,1,1,0,0,0,0,0]
> plt_fisher_2p,fishout
```

The final vector of 1's and 0's must be the same length as the number of parameters in Fisher matrix. This can be found using `print,n_elements(fish.sigma_marg)` or similar. To plot the 2D marginalised ellipse iCosmo uses a custom plotting routine called `plt_fisher_2p`. This actually creates a likelihood in the parameter space and then plots a filled contours at the 1-σ two parameter level.

A more elegant way (not creating the whole likelihood surface) of plotting Fisher ellipses is the following.

- First create a correlation matrix that is defined by $C_{ij} = F^{-1}_{ij}/(F^{-1}_{ii}F^{-1}_{jj})$.
- Define an angle $\phi = \cos^{-1}(C_{ij})$.
- Now trace out a line in x and y (by looping over θ) where $x = x_0 + \sqrt{(}2.3)\sqrt{(}F^{-1}_{ii})\sin(\theta)$ and $y = y_0 + \sqrt{(}2.3)\sqrt{(}F^{-1}_{jj})\sin(\theta - \phi)$.

- Plot the line (x,y).

This will create an ellipse showing the expected 1-σ two parameter contour for parameters i and j.