

# Project 1

## ***Linear Algebra, Optimization, Fisher Information Matrix, and MCMC***

This notebook is arranged in cells. Texts are usually written in the markdown cells, and here you can use html tags (make it bold, italic, colored, etc). You can double click on this cell to see the formatting.

The ellipsis (...) are provided where you are expected to write your solution but feel free to change the template (not over much) in case this style is not to your taste.

*Hit "Shift-Enter" on a code cell to evaluate it. Double click a Markdown cell to edit.*

*Write your partner's name here (if you have one).*

---

## Link Okpy

In [ ]:

```
from client.api.notebook import Notebook
ok = Notebook('project1.ok')
_ = ok.auth(inline = True)
```

## Imports

In [ ]:

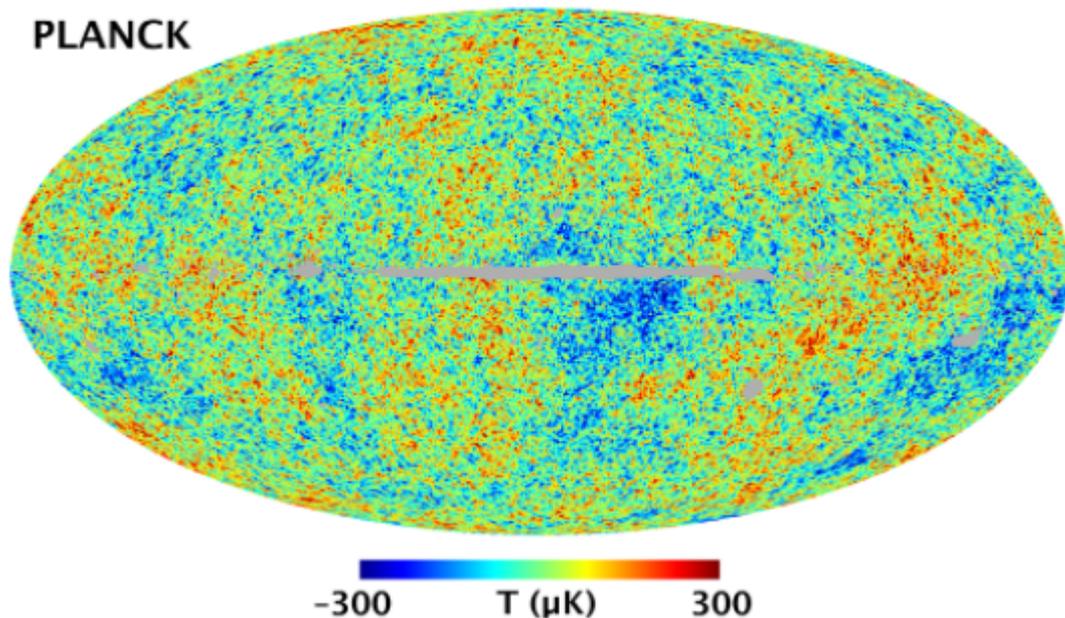
```
import numpy as np
from scipy.integrate import quad
#For plotting
import matplotlib.pyplot as plt
%matplotlib inline
```

---

## Problem 1 - Constraining the cosmological parameters using the Planck power spectrum

*Planck* is the third-generation space telescope, following COBE and WMAP, and it aims to determine the geometry and content of the Universe by observing the cosmic microwave background radiation (CMB), emitted around 380,000 years after the Big Bang. Permeating the whole universe and containing information on the properties of the early Universe, the CMB is widely known as the strongest evidence for the Big Bang model.

Measuring the spectrum of the CMB, we confirm that it is very close to the radiation from an ideal blackbody, and fluctuations in the spectrum are very small. Averaging over all locations, its mean temperature is  $2.725K$ , and its root mean square temperature fluctuation is  $\langle (\frac{\delta T}{T})^2 \rangle^{1/2} = 1.1 \times 10^{-5}$  (i.e. the temperature of the CMB varies by only  $\sim 30 \mu K$  across the sky).



Suppose you observe the fluctuations  $\delta T/T$ . Since we are taking measurements on the surface of a sphere, it is useful to expand  $\delta T/T$  in spherical harmonics (because they form a complete set of orthogonal functions on the sphere):

$$\frac{\delta T}{T}(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{lm} Y_{lm}(\theta, \phi)$$

In flat space, we can do a Fourier transform of a function  $f(x)$  as  $\sum_k a_k e^{ikx}$  where  $k$  is the wavenumber, and  $|a_k|$  determines the amplitude of the mode. For spherical harmonics, instead of  $k$ , we have  $l$ , the number of the modes along a meridian, and  $m$ , the number of modes along the equator. So  $l$  and  $m$  determine the wavelength ( $\lambda = 2\pi/l$ ) and shape of the mode, respectively.

In cosmology, we are mostly interested in learning the statistical properties of this map and how different physical effects influence different physical scales, so it is useful to define the correlation function  $C(\theta)$  and split the CMB map into different scales.

Suppose that we observe  $\delta T/T$  at two different points on the sky. Relative to an observer, they are in direction  $\hat{n}$  and  $\hat{n}'$  and are separated by an angle  $\theta$  given by  $\cos\theta = \hat{n} \cdot \hat{n}'$ . Then, we can find the correlation function by multiplying together the values of  $\delta T/T$  at the two points and average the product over all points separated by the angle  $\theta$ .

$$C(\theta)^{TT} = \left\langle \frac{\delta T}{T}(\hat{n}) \frac{\delta T}{T}(\hat{n}') \right\rangle_{\hat{n} \cdot \hat{n}' = \cos\theta}$$

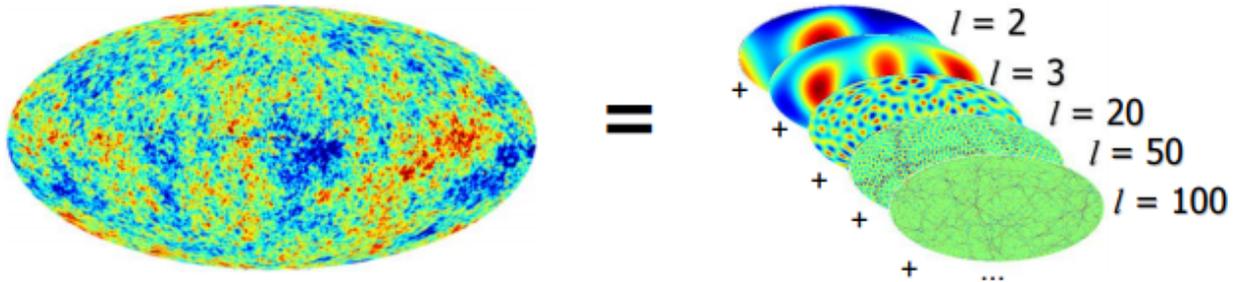
The above expression is specific to the temperature fluctuations, but we can also do a similar analysis for the polarization map of the CMB. (The CMB is polarized because it was scattered off of free electrons during decoupling.) We decompose the polarization pattern in the sky into a curl-free "E-mode" and grad-free "B-mode."

However, the CMB measurements (limited by the experiment resolution and the patch of sky examined) tell us about  $C(\theta)$  over only a limited range of angular scales. (i.e. the precise values of  $C(\theta)$  for all angles from  $\theta = 0$  to  $\theta = 180^\circ$  is not known.) Hence, using the expansion of  $\delta T/T$  in spherical harmonics, we write the correlation function as:

$$C(\theta) = \frac{1}{4\pi} \sum_{l=0}^{\infty} (2l+1) C_l P_l(\cos\theta)$$

where  $P_l$  are the Legendre polynomials.

So we break down the correlation function into its multipole moments  $C_l$ , which is the angular power spectrum of the CMB.



Remember that  $\lambda = 2\pi/l$ . So  $C_l$  measures the amplitude as a function of wavelength. (

$C_l = \frac{1}{2l+1} \sum_{m=-l}^l |a_{lm}|^2$ ). In this problem, we will consider the E-mode power spectrum

$$C_l^{EE} = \frac{1}{2l+1} \sum_{m=-l}^l |a_{lm}^E|^2$$

The CMB angular power spectrum is usually expressed in terms of  $D_l = l(l+1)C_l/2\pi$  (in unit of  $\mu K^2$ ) because this better shows the contribution toward the variance of the temperature fluctuations.

Cosmologists built a software called "cosmological boltzmann code" which computes the theoretical power spectrum given cosmological parameters, such as the Hubble constant and the baryon density. Therefore, we can fit the theory power spectrum to the measured one in order to obtain the best-fit parameters.

Here, we consider six selected cosmological parameters,

$\vec{\theta} = [\theta_1, \theta_2, \dots, \theta_6] = [H_0, \Omega_b h^2, \Omega_c h^2, n_s, A_s, \tau]$ . ( $H_0$  = Hubble constant,  $\Omega_b h^2$  = physical baryon density parameter,  $\Omega_c h^2$  = physical cold dark matter density parameter,  $n_s$  = scalar spectral index,  $A_s$  = curvature fluctuation amplitude,  $\tau$  = reionization optical depth.). We provide you with the measured CMB E-mode power spectrum from Planck Data Release 2. Then, assuming a simple linear model of the CMB power spectrum (i.e. assuming its response to those parameters are linear), we estimate the best-fit values of  $\vec{\theta}$  using linear algebra and Gauss-Newton optimization and plot their 1- $\sigma$ , 2- $\sigma$  confidence regions.

Then, how do we build a linear model of the theory power spectrum? Suppose a very simple scenario where you wish to determine the best-fit value of  $H_0$  assuming all the other parameters are already known (so

$\theta = H_0$  in this case). The measurements from WMAP (the CMB satellite which preceded Planck and consequently had a lower resolution) estimate that  $H_0 = 73.2$ . You take it as your starting value  $\theta_{ini}$  and compute the theory power spectrum there using the Boltzmann code. You also compute the derivative of  $D_l$  with respect to  $\theta$  at  $\theta_{ini}$ . Then, you can estimate the power spectrum as you perturb  $\theta$  around  $\theta_{ini}$ :

$$D_l^{model}(\theta = \theta_{ini} + \delta\theta) = D_l^{model}(\theta_{ini}) + \left. \frac{\partial D_l}{\partial \theta} \right|_{\theta=\theta_{ini}} \delta\theta.$$

From Planck, you get the measured  $D_l$  and error  $\sigma_l$ , so you can find the best-fit value of  $H_0$  which minimizes  $\chi^2 = ((D_l^{measured} - D_l^{model})/\sigma_l)^2$ . Also note that you expect the best-fit values from Planck will be close to WMAP estimate, so the above linear model is a valid approximation.

Now, we can similarly build a simple linear model power spectrum with six parameters. We take  $\vec{\theta}_{ini}$  as an estimate of the cosmological parameters from WMAP data

([https://lambda.gsfc.nasa.gov/product/map/dr2/params/lcdm\\_wmap.cfm](https://lambda.gsfc.nasa.gov/product/map/dr2/params/lcdm_wmap.cfm)

([https://lambda.gsfc.nasa.gov/product/map/dr2/params/lcdm\\_wmap.cfm](https://lambda.gsfc.nasa.gov/product/map/dr2/params/lcdm_wmap.cfm))).

$$D_l^{model}(\vec{\theta} = \vec{\theta}_{ini} + \delta\vec{\theta}) = D_l^{model}(\vec{\theta}_{ini}) + \sum_{i=1}^6 \left. \frac{\partial D_l}{\partial \theta_i} \right|_{\vec{\theta}=\vec{\theta}_{ini}} \delta\theta_i$$

So you can find the best-fit values of the above six cosmological parameters ( $\vec{\theta}_{best-fit}$ ) which minimizes

$$\chi^2(\vec{\theta}) = \sum_{l=l_{min}}^{l_{max}} \left( \frac{D_l^{measured} - D_l^{model}(\vec{\theta})}{\sigma_l} \right)^2$$

(i.e. when  $\vec{\theta} = \vec{\theta}_{best-fit}$ ,  $\chi^2$  is minimized.)

*References :*

Intro to Cosmology, Barbara Ryden

[http://folk.uio.no/hke/AST5220/v11/AST5220\\_2\\_2011.pdf](http://folk.uio.no/hke/AST5220/v11/AST5220_2_2011.pdf)

([http://folk.uio.no/hke/AST5220/v11/AST5220\\_2\\_2011.pdf](http://folk.uio.no/hke/AST5220/v11/AST5220_2_2011.pdf))

<http://cosmology.berkeley.edu/~yuki/CMBpol/> (<http://cosmology.berkeley.edu/~yuki/CMBpol/>)

The below cell defines  $l, D_l^{measured}, \sigma_l, \vec{\theta}_{ini}, D_l^{model}(\vec{\theta}_{ini}), \left. \frac{\partial D_l}{\partial \theta_i} \right|_{\vec{\theta}=\vec{\theta}_{ini}}$  (In problem 1, we only consider the CMB E-mode power spectrum, so  $D_l$  refers to  $D_l^{EE}$ .)

Here, we set  $l_{min} = 2, l_{max} = 2000$ , and we have 92  $l$ -bins in this range (For  $2 \leq l < 30$ , the power spectra are not binned ( $\Delta l = 1$ ), and for  $30 \leq l < 2000$ , they are binned, and the bin size is  $\Delta l = 30$ ). We obtain the measured and model power spectrum in that 92  $l$ -bins.

In [ ]:

```
# Load Data

# Measured power spectra from Planck
data = np.loadtxt("Project1_EE_measured.dat")
# l (same for all model and measured power spectrum)
ell = data[:,0]
#  $D_l^{EE}$  (measured)
EE_measured = data[:,1]
# and error
error_EE_measured = data[:,2]

# initial estimate of the parameters ( $\theta_{ini}$ ) - from https://lambda.gsfc.nasa.gov/product/map/dr2/params/lcdm_wmap.cfm
H0 = 73.2
ombh2 = 0.02229
omch2 = 0.1054
ns = 0.958
As = 2.347e-9
tau = 0.089

theta_ini = np.array([H0, ombh2, omch2, ns, As, tau])

# Model power spectra given  $\theta_{ini}$  (calculated at the same ell bins as the measured power spectrum)
data = np.loadtxt("Project1_EE_model_at_theta_ini.dat")
#  $D_l^{EE}$  (model)
EE_model = data[:,1]

# Derivative of the power spectra at  $\theta = \theta_{ini}$  (calculated at the same ell bins as the measured power spectrum)
data = np.loadtxt("Project1_derivative_EE_at_theta_ini.dat")
# Derivative of  $D_l^{EE}$  with respect to six parameters
# ([theta1, theta2, theta3, theta4, theta5, theta6] = [H_0, \Omega_b h^2, \Omega_c h^2, n_s, A_s, \tau])
deriv_DlEE_theta1 = data[:,1]
deriv_DlEE_theta2 = data[:,2]
deriv_DlEE_theta3 = data[:,3]
deriv_DlEE_theta4 = data[:,4]
deriv_DlEE_theta5 = data[:,5]
deriv_DlEE_theta6 = data[:,6]
```

1. Plot the measured power spectrum with errorbar. Also, plot the model power spectrum on top, by interpolating between the data points. You should find that the data from Planck does not fit to the model very well. To better see the low- $l$  measurements, also plot both spectra in the range  $2 \leq l < 30$ . Remember that the power spectra  $D_l$  have units of  $\mu K^2$ . Don't forget to label all plots.

In [ ]:

...

2. Using the techniques from linear algebra (normal equations, SVD, etc), find the best-fit cosmological parameters ( $\vec{\theta}_{best-fit}$ ). Print  $\vec{\theta}_{best-fit}$ .

Hint (only a suggestion):

$$\chi^2 = \sum_{l=l_{min}}^{l_{max}} \left( \frac{D_l^{measured} - D_l^{model}(\vec{\theta}_{ini}) - \sum_{j=1}^6 \left. \frac{\partial D_l}{\partial \theta_j} \right|_{\vec{\theta}=\vec{\theta}_{ini}} \delta \theta_j}{\sigma_l} \right)^2$$

where  $\delta \vec{\theta} = \vec{\theta} - \vec{\theta}_{ini}$ .

Note that the only variable in the above equation is  $\vec{\theta}$ . So we can re-write the above function as:

$$\chi^2 = \sum_{l=l_{min}}^{l_{max}} \left( \frac{D_l^{measured} - D_l^{model}(\vec{\theta}_{ini}) + \sum_{j=1}^6 \left. \frac{\partial D_l}{\partial \theta_j} \right|_{\vec{\theta}=\vec{\theta}_{ini}} (\vec{\theta}_{ini})_j - \sum_{j=1}^6 \left. \frac{\partial D_l}{\partial \theta_j} \right|_{\vec{\theta}=\vec{\theta}_{ini}} \theta_j}{\sigma_l} \right)^2$$

Now let  $y_l = D_l^{measured} - D_l^{model}(\vec{\theta}_{ini}) + \sum_{j=1}^6 \left. \frac{\partial D_l}{\partial \theta_j} \right|_{\vec{\theta}=\vec{\theta}_{ini}} (\vec{\theta}_{ini})_j$  (Note:  $y_l$  is independent of  $\vec{\theta}$ ).

Then, we can simplify the above  $\chi^2$  function as:

$$\chi^2 = \sum_{l=l_{min}}^{l_{max}} \left( \frac{y_l - \sum_{j=1}^6 \left. \frac{\partial D_l}{\partial \theta_j} \right|_{\vec{\theta}=\vec{\theta}_{ini}} \theta_j}{\sigma_l} \right)^2$$

Here, we are trying to determine  $\theta_1, \dots, \theta_6$ . Remember how we solved the linear least squares problem using the normal equations. Re-write the  $\chi^2$  function in a matrix form and find  $\vec{\theta}$  which minimizes  $\chi^2$ .

In [ ]:

...

We can define  $\chi^2(\vec{\theta}) = \vec{r}(\vec{\theta})^T \vec{r}(\vec{\theta})$  where  $\vec{r} = \frac{D_l^{model}(\vec{\theta}) - D_l^{measured}}{\sigma_l}$ . (so  $\vec{r}$  is a vector of length 92.)

We then compute the gradient and the Hessian of  $\chi^2$  to apply the Gauss-Newton method. (This is a linear least squares problem, so using the Gauss-Newton method is equivalent to using the normal equations in this case - yes, you may think that it is silly to use the Gauss-Newton here! So we are expected to reach minimum just after one iteration. The Jacobian  $J$  is identical to the design matrix  $A$ .)

The  $j$ th component of the gradient is:

$$(\nabla \chi^2(\vec{\theta}))_j = 2 \sum_l r_l(\vec{\theta}) \frac{\partial r_l}{\partial \theta_j}(\vec{\theta})$$

where  $\frac{\partial r_l}{\partial \theta_j} = \frac{1}{\sigma_l} \frac{\partial D_l}{\partial \theta_j} \Big|_{\vec{\theta}=\vec{\theta}_{ini}}$

Now, the Jacobian matrix  $J(\vec{\theta})$  is:  $\begin{bmatrix} \frac{\partial r_{\{min\}}}{\partial \theta_1}(\vec{\theta}) & \dots & \frac{\partial r_{\{min\}}}{\partial \theta_6}(\vec{\theta}) \\ \dots & \dots & \dots \\ \frac{\partial r_{\{max\}}}{\partial \theta_1}(\vec{\theta}) & \dots & \frac{\partial r_{\{max\}}}{\partial \theta_6}(\vec{\theta}) \end{bmatrix}$

The gradient of  $\chi^2$  can be written as:

$$\nabla \chi^2(\vec{\theta}) = 2J(\vec{\theta})^T \vec{r}(\vec{\theta})$$

Similarly, the  $(i, j)$ th component of the Hessian matrix of  $\chi^2$  is given by:

$$\frac{\partial^2(\chi^2)}{\partial \theta_i \partial \theta_j}(\vec{\theta}) = 2 \sum_l \frac{\partial r_l}{\partial \theta_i}(\vec{\theta}) \frac{\partial r_l}{\partial \theta_j}(\vec{\theta})$$

(Here,  $\frac{\partial^2 r_l}{\partial \theta_i \partial \theta_j} = 0$ )

Because our model power spectrum is linear, we can write the Hessian matrix simply as

$$H(\vec{\theta}) = 2J(\vec{\theta})^T J(\vec{\theta}).$$

Then, using Newton's method, we can find  $\vec{\theta}$  which minimizes  $\chi^2$ :

$$\vec{\theta}^{(k+1)} = \vec{\theta}^{(k)} - (J(\vec{\theta}^{(k)})^T J(\vec{\theta}^{(k)}))^{-1} J(\vec{\theta}^{(k)})^T \vec{r}(\vec{\theta}^{(k)})$$

We have a simple linear model in this case, so we are expected to reach the minimum after one step.

3. Using the Gauss-Newton optimization, find the best-fit parameters ( $\vec{\theta}_{best-fit}$ ). Iterate until you reach the minimum (Show that you get the best-fit values after one step). Does your result agree with Part 2?

In [ ]:

...

We can compute the covariance matrix as  $(J(\vec{\theta})^T J(\vec{\theta}))^{-1}$ . (Remember that the covariance matrix in the normal equations is  $(A^T A)^{-1}$  where  $A$  is the design matrix.) From this, we can plot 1-d and 2-d constraints on the parameters. (See Fig. 6 in Planck 2015 paper <https://arxiv.org/pdf/1502.01589v3.pdf> (<https://arxiv.org/pdf/1502.01589v3.pdf>))

**1-d constraint** (corresponding to the plots along the diagonal in Fig. 6, Planck 2015 paper):

First, the  $i$ th diagonal element of the covariance matrix correspond to  $\sigma(\theta_i)^2$ . Then, we can plot 1-d constraints on the parameter  $\theta_i$  assuming a normal distribution with mean =  $(\vec{\theta}_{best-fit})_i$  and variance =  $\sigma(\theta_i)^2$ .

**2-d constraint** (off-diagonal plots in Fig. 6, Planck 2015 paper):

Consider two parameters  $\theta_i$  and  $\theta_j$  from  $\vec{\theta}$ . Now marginalize over other parameters - in order to marginalize over other parameters, you can simply remove those parameters' row and column from the full covariance matrix. (i.e. From the full covariance matrix, you know the variance of all six parameters and their covariances with each other. So build a smaller dimension - 2 x 2 - covariance matrix from this.) - and obtain a  $2 \times 2$  covariance matrix:

$$C_{ij} = \begin{pmatrix} \sigma(\theta_i)^2 & \text{Cov}(\theta_i, \theta_j) \\ \text{Cov}(\theta_i, \theta_j) & \sigma(\theta_j)^2 \end{pmatrix}$$

Now, we can plot the 2-dimensional confidence region ellipses from this matrix. The lengths of the ellipse axes are the square root of the eigenvalues of the covariance matrix, and we can calculate the counter-clockwise rotation of the ellipse with the rotation angle:

$$\phi = \frac{1}{2} \arctan\left(\frac{2 \cdot \text{Cov}(\theta_i, \theta_j)}{\sigma(\theta_i)^2 - \sigma(\theta_j)^2}\right) = \arctan\left(\frac{\vec{v}_1(y)}{\vec{v}_1(x)}\right)$$

where  $\vec{v}_1$  is the eigenvector with the largest eigenvalue. So we calculate the angle of the largest eigenvector towards the x-axis to obtain the orientation of the ellipse.

Then, we multiply the axis lengths by some factor depending on the confidence level we are interested in. For 68%, this scale factor is  $\sqrt{\Delta\chi^2} \approx 1.52$ . For 95%, it is  $\sqrt{\Delta\chi^2} \approx 2.48$ .

4. Plot 1-d and 2-d constraints on the parameters. For 2-d plot, show 68% and 95% confidence ellipses for each pair of parameters. You can arrange those subplots in a triangle shape, as in Fig. 6, Planck 2015 (<https://arxiv.org/pdf/1502.01589v3.pdf>).

Hint: For plotting ellipses, see HW4 Q2-7 solution (<http://datahub.berkeley.edu/user-redirect/interact?account=bccp&repo=seljak-phy151-fall-2017&branch=master&path=Homework/HW4/HW4-solution.ipynb> (<http://datahub.berkeley.edu/user-redirect/interact?account=bccp&repo=seljak-phy151-fall-2017&branch=master&path=Homework/HW4/HW4-solution.ipynb>))

In [ ]:

...

5. Plot  $D_l^{model}(\vec{\theta} = \vec{\theta}_{ini})$  and  $D_l^{model}(\vec{\theta} = \vec{\theta}_{best-fit})$  as well as  $D_l^{measured}$  with errorbar. Show that with the best-fit parameters you obtained, the model power spectrum fits better to the measured data.

In [ ]:

...

## Problem 2 - Fisher prediction for future CMB surveys

In class, we learned that the Fisher information matrix is useful for designing an experiment; we can vary the experiment design and predict the level of the expected error on any given parameter. In this problem, we aim to determine how well a low-noise, high-resolution future CMB survey would do in constraining the cosmological parameters.

The Fisher matrix is defined as the ensemble average of the Hessian of the log-likelihood ( $\ln \mathcal{L}$ ) with respect to the given parameters  $\vec{\theta}$ :

$$F_{ij} = - \left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial \theta_i \partial \theta_j} \right\rangle$$

Here we take the model CMB power spectrum as our observables. (Here we consider the auto-correlations  $D_l^{TT}$ ,  $D_l^{EE}$  and cross-correlation  $D_l^{TE}$  obtained from the boltzmann code using the best-fit cosmological parameters from Planck, <https://arxiv.org/pdf/1502.01589v3.pdf> (<https://arxiv.org/pdf/1502.01589v3.pdf>)). Then, we can estimate the Fisher matrix between two parameters  $\theta_i$  and  $\theta_j$  as:

$$F_{ij} = \sum_l \sum_k \frac{1}{(\sigma_l^k)^2} \frac{\partial D_l^k}{\partial \theta_i} \frac{\partial D_l^k}{\partial \theta_j}$$

where we sum over the CMB auto- and cross-power spectra  $D_l^k = [D_l^{TT}, D_l^{EE}, D_l^{TE}]$ , and we assume that there is no correlation between them.  $\sigma^2$  is the variance of  $D_l$  and noise:

$$(\sigma_l^k)^2 = \frac{2}{(2l+1) \cdot f_{sky} \cdot \Delta l} (D_l^k + N_l^k)^2$$

where  $f_{sky}$  is the fraction of the sky covered by the survey. Assume that  $f_{sky} = 1$  for the sake of simplicity.  $\Delta l$  is the size of  $l$ -bin. (Remember that for  $2 \leq l < 30$ , the power spectra are not binned ( $\Delta l = 1$ ), and for  $30 \leq l < 2000$ , they are binned, and the bin size is  $\Delta l = 30$ .) In this problem, first take the noise from Planck. We provide you with  $\sigma_l$  for  $D_l^{TT}$ ,  $D_l^{EE}$ ,  $D_l^{TE}$  from Planck.

Then, assume that we have an ideal, zero-noise CMB survey with  $N_l = 0$ . However, we are still intrinsically limited on the number of independent modes we can measure (there are only  $(2l+1)$  of them) -

$C_l = \frac{1}{2l+1} \sum_{m=-l}^l \langle |a_{lm}|^2 \rangle$ . This leads that we get an intrinsic error (called "cosmic variance") in our estimate of  $C_l$ . So we approximate that

$$(\sigma_l^{EE})^2 = \frac{2}{(2l+1) \cdot f_{sky} \cdot \Delta l} (D_l^{EE})^2, \quad (\sigma_l^{TT})^2 = \frac{2}{(2l+1) \cdot f_{sky} \cdot \Delta l} (D_l^{TT})^2,$$

$$(\sigma_l^{TE})^2 = \frac{2}{(2l+1) \cdot f_{sky} \cdot \Delta l} \frac{(D_l^{TE})^2 + D_l^{TT} D_l^{EE}}{2}$$

Finally, we can obtain the covariance matrix  $C$  by inverting the Fisher matrix  $F$ :

$$[C] = [F]^{-1}$$

*References:*

Fisher Matrix Forecasting Review, Nicholas Kern

<https://arxiv.org/pdf/0906.4123.pdf> (<https://arxiv.org/pdf/0906.4123.pdf>)

1. First, load the measurement errors ( $\sigma_l^{TT}$ ,  $\sigma_l^{EE}$ ,  $\sigma_l^{TE}$ ), model power spectrum ( $D_l^{TT}$ ,  $D_l^{EE}$ ,  $D_l^{TE}$ ) and their derivatives with respect to six cosmological parameters evaluated at the best-fit values from Planck ( $\frac{\partial D_l^{TT}}{\partial H_0} \Big|_{\vec{\theta}=\vec{\theta}_{best-fit}}$ ,  $\frac{\partial D_l^{TT}}{\partial \Omega_b h^2} \Big|_{\vec{\theta}=\vec{\theta}_{best-fit}}$ , etc). With the measurement errors from Planck, construct the Fisher matrix and the covariance matrix (you can use the `numpy.linalg.inv` for the matrix inversion). Evaluate the constraints on six parameters  $\sigma(H_0)$ ,  $\sigma(\Omega_b h^2)$ ,  $\dots$ ,  $\sigma(\tau)$  (corresponding to the square root of the diagonal entries of the covariance matrix). Print the results.

In [ ]:

```
# Load data

# Best-fit values of the cosmological parameters from https://arxiv.org/pdf/150
2.01589v3.pdf
H0      = 67.27
ombh2   = 0.02225
omch2   = 0.1198
ns      = 0.9645
As      = 2.2065e-9
tau     = 0.079

theta_best_Planck = np.array([H0, ombh2, omch2, ns, As, tau])

# Planck noise

# sigma_l for D_l^EE
data = np.loadtxt("Project1_EE_measured.dat")
# l (same for all power spectrum)
ell = data[:,0]
# and error
error_EE = data[:,2]

# sigma_l for D_l^TT
data = np.loadtxt("Project1_TT_measured.dat")
# and error
error_TT = data[:,2]

# sigma_l for D_l^TE
data = np.loadtxt("Project1_TE_measured.dat")
# and error
error_TE = data[:,2]

# Model power spectra given theta_best_Planck (calculated at the same ell bins as
the measured power spectrum)

# D_l^EE (model)
data = np.loadtxt("Project1_EE_model_at_theta_best_Planck.dat")
EE_model_Planck = data[:,1]

# D_l^TT (model)
data = np.loadtxt("Project1_TT_model_at_theta_best_Planck.dat")
TT_model_Planck = data[:,1]

# D_l^TE (model)
data = np.loadtxt("Project1_TE_model_at_theta_best_Planck.dat")
TE_model_Planck = data[:,1]

# Derivative of the power spectrum given theta_best_Planck (calculated at the same
ell bins as the measured power spectrum)

# Derivative of D_l^EE with respect to six parameters
# ([theta1, theta2, theta3, theta4, theta5, theta6] = [H_0, \Omega_b h^2, \Omega_c
h^2, n_s, A_s, \tau])
data = np.loadtxt("Project1_derivative_EE_at_theta_best_Planck.dat")
deriv_DlEE_theta1 = data[:,1]
deriv_DlEE_theta2 = data[:,2]
```

```

deriv_DEE1_theta3 = data[:,3]
deriv_DLEE_theta4 = data[:,4]
deriv_DLEE_theta5 = data[:,5]
deriv_DLEE_theta6 = data[:,6]

# Derivative of D_l^TT with respect to six parameters
data = np.loadtxt("Project1_derivative_TT_at_theta_best_Planck.dat")
deriv_DlTT_theta1 = data[:,1]
deriv_DlTT_theta2 = data[:,2]
deriv_DlTT_theta3 = data[:,3]
deriv_DlTT_theta4 = data[:,4]
deriv_DlTT_theta5 = data[:,5]
deriv_DlTT_theta6 = data[:,6]

# Derivative of D_l^TE with respect to six parameters
data = np.loadtxt("Project1_derivative_TE_at_theta_best_Planck.dat")
deriv_DlTE_theta1 = data[:,1]
deriv_DlTE_theta2 = data[:,2]
deriv_DlTE_theta3 = data[:,3]
deriv_DlTE_theta4 = data[:,4]
deriv_DlTE_theta5 = data[:,5]
deriv_DlTE_theta6 = data[:,6]

```

In [ ]:

...

2. Same as in Problem1-Part4. From the covariance matrix, plot 1-d and 2-d constraints on the parameters. Note that the best-fit values of six parameters are already given in Part 1 (we just use the values from the Planck paper). For 2-d plot, show 68% and 95% confidence ellipses for pairs of parameters. You can arrange those subplots in a triangle shape, as in Fig. 6, Planck 2015 (<https://arxiv.org/pdf/1502.01589v3.pdf>).

In [ ]:

...

3. Repeat Part 1 and 2 assuming  $N_l^k = 0$ . (How well does a zero-noise CMB survey constrain the cosmological parameters?)

In [ ]:

...

4. Combine Part 2 and Part 3 and compare. (First plot your results from Part 2 (1-d and 2-d constraints using the Planck power spectra and noise. Then, plot Part 3 results (assuming zero noise) on top with different colors. Note that your 1-d constraints in Part 3 are more sharply peaked Gaussians (with much smaller variances), so you can scale them so that its peak amplitudes match with your results from Part 2.)

In [ ]:

...

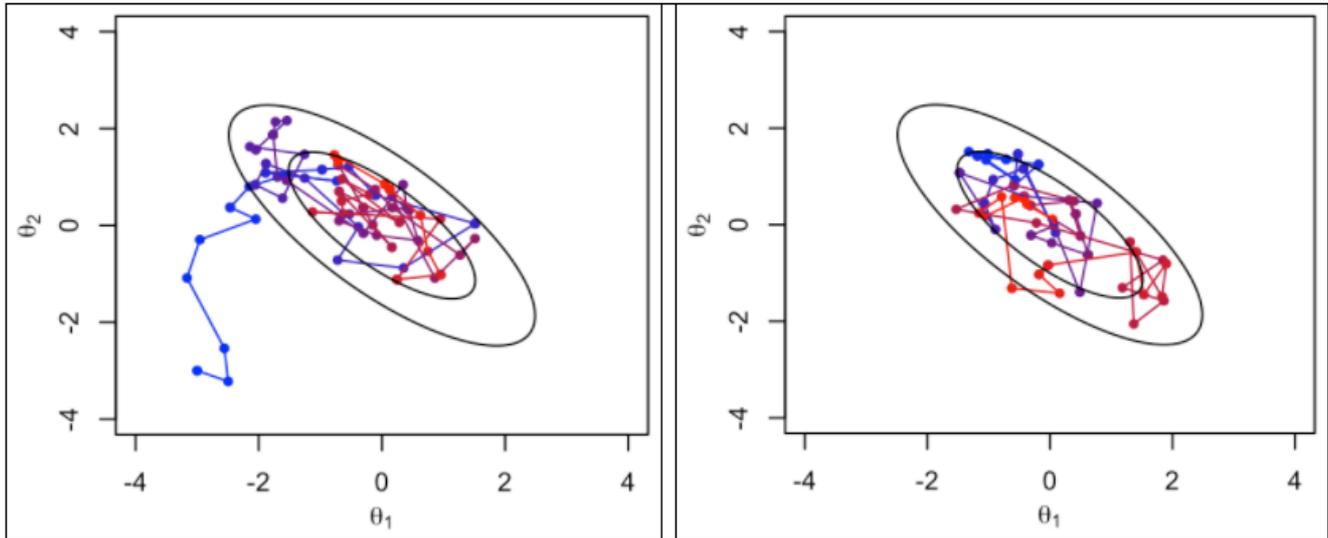
5. In Problem1-Part4, you estimated the best-fit values of the cosmological parameters and their 1, 2- $\sigma$  confidence regions. In Problem2-Part2, starting from the best-fit values from the Planck 2015 paper, you constrained six cosmological parameters assuming that you have a zero-noise future CMB survey. Compare your results with Table 3 and Figure 6 in <https://arxiv.org/pdf/1502.01589v3.pdf>.

Answer:

---

### Problem 3 - Planck MCMC chain

Markov chain Monte Carlo is a general method based on drawing values of  $\theta$  from approximate distributions and then correcting those draws to better approximate the target posterior distribution. The sampling is done sequentially, with the distribution of the sampled draws depending on the last value drawn - hence, the draws from a Markov chain. (p. 275, *Bayesian Data Analysis*, Andrew Gelman et al.) (Remember that a sequence  $x_1, x_2, \dots$  of random events is called a Markov chain if  $x_{n+1}$  depends explicitly on  $x_n$  only (and not explicitly on previous steps).) Here, we consider six selected cosmological parameters: [ $H_0, \Omega_b h^2, \Omega_c h^2, n_s, A_s, \tau$ ], so the "chain" in this case is a random walk through the parameter space.



from <https://github.com/KIPAC/StatisticalMethods/blob/master/chunks/montecarlo1.ipynb>  
(<https://github.com/KIPAC/StatisticalMethods/blob/master/chunks/montecarlo1.ipynb>)

As shown in the above figure, chains take time to converge to the target distribution, and you can determine the "burn-in" period, the number of sequences it takes to reach convergence.

In this problem, we provide you MCMC chains (using Planck low and high- $l$  temperature data with lensing reconstruction) from Planck Data Release 1 ([http://irsa.ipac.caltech.edu/data/Planck/release\\_1/ancillary-data/](http://irsa.ipac.caltech.edu/data/Planck/release_1/ancillary-data/) ([http://irsa.ipac.caltech.edu/data/Planck/release\\_1/ancillary-data/](http://irsa.ipac.caltech.edu/data/Planck/release_1/ancillary-data/))). You can plot the chains in the parameter space and estimate the posterior distribution.

#### References:

Bayesian Data Analysis, Andrew Gelman et al.

<https://github.com/KIPAC/StatisticalMethods/blob/master/chunks/montecarlo1.ipynb>  
(<https://github.com/KIPAC/StatisticalMethods/blob/master/chunks/montecarlo1.ipynb>)

1. First, we give you one Planck chain without removing the burn-in. In this case, the parameter space is ( $H_0, \Omega_b h^2$ ). Estimate the burn-in period.

In [ ]:

```
# Load data
data = np.loadtxt("Planck_chain_with_burnin.txt")
# H0
theta1 = data[:,23]
# Omega_b h^2
theta2 = data[:,2]

# Plot chain
plt.plot(theta1, theta2, 'x-')
plt.xlabel('$H_0$')
plt.ylabel('$\Omega_b h^2$')
plt.show()
```

Answer:

2. Now, we provide you with 8 independent Planck MCMC chains. For each chain, we load the data for six cosmological parameters we are considering,  $[H_0, \Omega_b h^2, \Omega_c h^2, n_s, A_s, \tau]$ . From the chain, estimate the posterior distribution of each of the six parameters. (Plot the 1-d posterior distribution and estimate its mean + standard deviation.)

In [ ]:

```
# Load data

# H_0
theta1_chain = np.zeros(shape=(8,1981))
# Omega_b h^2
theta2_chain = np.zeros(shape=(8,1981))
# Omega_c h^2
theta3_chain = np.zeros(shape=(8,1981))
# n_s
theta4_chain = np.zeros(shape=(8,1981))
# A_s
theta5_chain = np.zeros(shape=(8,1981))
# tau
theta6_chain = np.zeros(shape=(8,1981))

# 8 Planck chains, each of length 1981 (so theta6_chain[1] contains values of tau in Planck chain 1)
for i in range(8):
    data = np.loadtxt("base_planck_lowl_post_lensing_%d.txt" %(i+1))
    theta1_chain[i] = data[:,27][0:1981]
    theta2_chain[i] = data[:,2][0:1981]
    theta3_chain[i] = data[:,3][0:1981]
    theta4_chain[i] = data[:,6][0:1981]
    theta5_chain[i] = data[:,29][0:1981]*1.e-9
    theta6_chain[i] = data[:,5][0:1981]
```

In [ ]:

...

3. For all pairs of the parameters, compute the covariance. Make a 2-d scatterplot of the chains (as in Problem3-Part 1). Then, plot 68% and 95% confidence ellipses on top of the scatterplots, as in Problem1-Part4. Compare your answers with Problem1-Part4 and Problem2-Part2.

In [ ]:

...

In MCMC, we need to make sure that chains converge to the posterior distribution. One useful test for convergence is "Gelman-Rubin statistic." For a given parameter,  $\theta$ , the  $R$  statistic compares the variance across chains with the variance within a chain. Intuitively, if the chains are random-walking in very different places, i.e. not sampling the same distribution,  $R$  will be large.

In detail, given chains  $J = 1, \dots, m$ , each of length  $n$ ,

Let  $B = \frac{n}{m-1} \sum_j (\bar{\theta}_j - \bar{\theta})^2$ , where  $\bar{\theta}_j$  is the average  $\theta$  for chain  $j$  and  $\bar{\theta}$  is the global average. This is proportional to the variance of the individual-chain averages for  $\theta$ .

Let  $W = \frac{1}{m} \sum_j s_j^2$ , where  $s_j^2$  is the estimated variance of  $\theta$  within chain  $j$ . This is the average of the individual-chain variances for  $\theta$ .

Let  $V = \frac{n-1}{n} W + \frac{1}{n} B$ . This is an estimate for the overall variance of  $\theta$ .

Finally,  $R = \sqrt{\frac{V}{W}}$ . We'd like to see  $R \approx 1$  (e.g.  $R < 1.1$  is often used). Note that this calculation can also be used to track convergence of combinations of parameters, or anything else derived from them.

Reference: <https://github.com/KIPAC/StatisticalMethods/blob/master/chunks/montecarlo1.ipynb> (<https://github.com/KIPAC/StatisticalMethods/blob/master/chunks/montecarlo1.ipynb>)

4. For all six parameters, compute  $R$  and determine if the condition  $R < 1.1$  is satisfied.

In [ ]:

...

The autocorrelation of a sequence, as a function of lag,  $k$ , is defined thusly:

$$\rho_k = \frac{\sum_{i=1}^{n-k} (\theta_i - \bar{\theta}) (\theta_{i+k} - \bar{\theta})}{\sum_{i=1}^{n-k} (\theta_i - \bar{\theta})^2} = \frac{\text{Cov}_i(\theta_i, \theta_{i+k})}{\text{Var}(\theta)}$$

The larger lag one needs to get a small autocorrelation, the less informative individual samples are.

5. Using `autocorrelation_plot` from `pandas` (<https://pandas.pydata.org/pandas-docs/stable/visualization.html#visualization-autocorrelation>), plot the auto-correlation of six parameters and determine that it gets small for large lag. The given Planck MCMC chains are already heavily thinned, so you will not see much autocorrelation.

In [ ]:

```
from pandas.tools.plotting import autocorrelation_plot
...
```

From the WMAP measurements, the best-fit value of  $\tau$  was  $0.089 \pm 0.013$ . However, as discussed in Komatsu et al. (2009), uncertainties in modelling polarized foreground emission are comparable to the statistical error in the WMAP  $\tau$  measurement. In particular, at the time of the WMAP9 analysis there was very little information available on polarized dust emission. So in the Planck analysis, cosmologists cleaned the WMAP maps for polarized dust emission, and this lowered  $\tau$  by  $1\sigma$  to  $\tau = 0.075 \pm 0.013$ .

(<https://arxiv.org/pdf/1502.01589.pdf> (<https://arxiv.org/pdf/1502.01589.pdf>))

6. From the Planck MCMC chain, we determined the mean value of the  $\tau$  posterior distribution. Now do the importance sampling by lowering  $\tau$  by  $1\sigma$ . What happens to the posterior?

In [ ]:

## To Submit

Execute the following cell to submit. If you make changes, execute the cell again to resubmit the final copy of the notebook, they do not get updated automatically.

**We recommend that all the above cells should be executed (their output visible) in the notebook at the time of submission.**

Only the final submission before the deadline will be graded.

In [ ]:

```
_ = ok.submit()
```