(DOD) 11.3.2 optimal quadratic estimator

We would like to find the maximum likelihood. As we know, this is equivalent to looking for the maximum log-likelihood

$$L = \ln \mathrm{prob}(\vec{\Delta} | C)$$

Please note that for a uniform prior, this is of course equivalent to searching for the maximum posterior prob$(\vec{\lambda} | \vec{\Delta})$ for some model parameter $\vec{\lambda}$.
Anyway. We will use the Newton-Raphson method to find our maximum:

$$\left.\frac{\partial L}{\partial \lambda}\right|_{\lambda = \bar{\lambda}} = 0$$

where $\bar{\lambda}$ is the parameter value λ maximizing the likelihood.

Let's start at some trial point λ^0 and expand $L = \ln \mathrm{prob}(\vec{\Delta} | C)$ around this

$$0 = L_{,\lambda_i}(\bar{\lambda}) = L_{,i}(\lambda^0) + \underbrace{L_{,ij}}_{\equiv L^{(2)}_{ij}}(\lambda^0)\,[\bar{\lambda} - \lambda^0]_j + \dots$$

$$\Rightarrow \quad [\bar{\lambda} - \lambda^0]_j \approx -(L^{(2)})^{-1}_{ji} L_{,i}$$

Let us denote $L_{,i} \equiv \frac{\partial L}{\partial \lambda_i}$ etc

$$\Rightarrow \quad \bar{\lambda}_j \approx \lambda^0_j - (L^{(2)})^{-1}_{ji} L_{,i}$$

where "$\approx$" becomes "$=$" in case that L is quadratic, i.e. gaussian likelihood. So quite a good guess.
Of course, iterate further to find maximum.

Let's consider our CMB Likelihood

$$\text{prob}(\vec{\Delta}|C) = \frac{(2\pi)^{-N_p/2}}{\sqrt{\det C}} \exp\left[-\frac{1}{2}\vec{\Delta}^T C^{-1} \vec{\Delta}\right]$$

Which you can think of as a prototype of a gaussian likelihood.

$$L = \ln \text{prob}(\vec{\Delta}|C) = \text{const} - \frac{1}{2}\ln\det C - \frac{1}{2}\vec{\Delta}^T C^{-1}\vec{\Delta}$$

$$= \text{const} - \frac{1}{2}\,\text{tr}\ln C - \frac{1}{2}\vec{\Delta}^T C^{-1}\vec{\Delta}$$

Where we used $\ln\det A = \text{tr}\ln A$.
Remember also that $\frac{d}{d\lambda}A^{-1} = -A^{-1}A_{,\lambda}A^{-1}$, then:

$$\frac{d}{d\lambda}\ln A = A^{-1}A_{,\lambda}$$

$$L_{,\lambda_i} = -\frac{1}{2}\,\text{Tr}\, C^{-1}C_{,i} + \frac{1}{2}\vec{\Delta}^T C^{-1}C_{,i}C^{-1}\vec{\Delta}$$

$$L_{,\lambda_i\lambda_j} = \frac{1}{2}\,\text{Tr}\, C^{-1}C_{,i}C^{-1}C_{,j} - \frac{1}{2}\,\text{Tr}\, C^{-1}C_{,ij} + \frac{1}{2}\vec{\Delta}^T C^{-1}C_{,ij}C^{-1}\vec{\Delta}$$

$$- \frac{1}{2}\vec{\Delta}^T C^{-1}C_{,i}C^{-1}C_{,j}C^{-1}\vec{\Delta} - \frac{1}{2}\vec{\Delta}^T C^{-1}C_{,i}C^{-1}C_{,j}C^{-1}\vec{\Delta}$$

$$= \frac{1}{2}\,\text{Tr}\, C^{-1}C_{,i}C^{-1}C_{,j} - \frac{1}{2}\,\text{Tr}\, C^{-1}C_{,ij} + \frac{1}{2}\vec{\Delta}^T C^{-1}C_{,ij}C^{-1}\vec{\Delta}$$

$$- \vec{\Delta}^T C^{-1}C_{,i}C^{-1}C_{,j}C^{-1}\vec{\Delta}$$

A few hours ago, I told you that the Fisher information matrix is the expectation value of minus the logarithm of $\text{prob}(\vec{\lambda}|\vec{\Delta})$. We follow Dodelson who implicitly assumes a flat prior (like we did) and hence

$$F_{ij} = -\langle L_{,ij} \rangle$$

Please note that for any matrix A,

$$\langle \vec{\Delta}^T A \vec{\Delta} \rangle = \langle \Delta_i A_{ij} \Delta_j \rangle = A_{ij} \langle \Delta_i \Delta_j \rangle$$
$$= A_{ij} C_{ij} = A_{ij} C_{ji} = \mathrm{Tr}\, AC \qquad ; C \text{ symmetric}$$

So the Fisher matrix is:

$$F_{ij} = -\frac{1}{2} \mathrm{tr}\, C^{-1} C_{,i} C^{-1} C_{,j} + \frac{1}{2} \mathrm{tr}\, C^{-1} C_{,ij} - \frac{1}{2} \vec{\Delta}^T C^{-1} C_{,ij} C^{-1} \vec{\Delta}$$
$$+ \langle \vec{\Delta}^T C^{-1} C_{,i} C^{-1} C_{,j} C^{-1} \vec{\Delta} \rangle$$

$$= -\frac{1}{2} \mathrm{tr}\, C^{-1} C_{,i} C^{-1} C_{,j} + \frac{1}{2} \mathrm{tr}\, C^{-1} C_{,ij} - \frac{1}{2} \mathrm{tr}\, C^{-1} C_{,ij} C^{-1} C$$
$$+ \mathrm{tr}\, C^{-1} C_{,i} C^{-1} C_{,j} C^{-1} C$$

$$= \frac{1}{2} \mathrm{tr}\, C^{-1} C_{,i} C^{-1} C_{,j}$$

Which we promised to proof a few hours ago.

As far as our maximum likelihood estimator is concerned, what one does is to use

$\langle L^{(2)} \rangle^{-1}_{ij}$ instead of $\left[(L^{(2)})^{-1} \right]_{ij}$ itself, so one gets

$$\hat{\lambda}_j = \lambda_j^0 + \frac{1}{2} (F^{-1})_{ji} \left\{ \vec{\Delta}^T C^{-1} C_{,i} C^{-1} \vec{\Delta} - \mathrm{Tr}\, C^{-1} C_{,i} \right\}$$

Start
11.12.06.

Our estimator $\hat{\lambda}_i$ is quadratic, i.e. of the general form $A\Delta^2+B$.

It is best used ~~the...~~ iteratively to find the best fitting parameters. In practice, only a few iterations are necessary. So we know the maximum likelihood point. But what about the errors on the parameters for these values? Think of $\hat{\lambda}$ as an estimator for the true parameter values and study its distribution. We know the distribution for signal and noise C_S and C_N which are both gaussian.

First, let's compute the expectation value

$$\langle \hat{\lambda}_i \rangle = \lambda_i^0 + \frac{1}{2}(F^{-1})_{ji}\left\{\langle \vec{\Delta}^T C^{-1} C_{,i} C^{-1} \vec{\Delta}\rangle - \mathrm{Tr}\, C^{-1} C_{,i}\right\}$$

C; $C_{,i}$, F_{ii} are all evaluated at trial point $\vec{\lambda}^0$. But $\langle \vec{\Delta}\vec{\Delta}^T\rangle$ is equal to the true covariance matrix $C(\bar{\lambda})$. Let's expand C about $\vec{\lambda}^{(0)}$:

$$C(\bar{\lambda}) = C(\lambda^0) + C_{,k}(\bar{\lambda}_k - \lambda_k^0)$$

$$\Rightarrow \langle \hat{\lambda}_i \rangle = \lambda_i^0 + \frac{1}{2}(F^{-1})_{ji}\left\{\mathrm{Tr}\, C^{-1} C_{,i} C^{-1} C(\bar{\lambda}) - \mathrm{Tr}\, C^{-1} C_{,i}\right\}$$

$$= \lambda_i^0 + \frac{1}{2}(F^{-1})_{ji}\left\{\mathrm{Tr}\, C^{-1} C_{,i} \cancel{C^{-1}} C + \mathrm{Tr}\, C^{-1} C_{,i} C^{-1} C_{,k}(\bar{\lambda}_k - \lambda_k^0) - \mathrm{Tr}\, C^{-1} C_{,i}\right\}$$

$$= \lambda_i^0 + \frac{1}{2}(F^{-1})_{ji}\left\{\bar{\lambda}_k - \lambda_k^0\right\} \underbrace{\mathrm{Tr}\, C^{-1} C_{,i} C^{-1} C_{,k}}_{=2F_{ik}}$$

$$= \lambda_i^0 + (\bar{\lambda}_k - \lambda_k^0)(F^{-1})_{ij} F_{ik} = \bar{\lambda}_j$$

So $\langle \hat{\lambda}_i \rangle = \bar{\lambda}_i$ which means that our estimator is unbiased ! The expectation values of $\hat{\lambda}_i$ are equal to the true parameters irrespective of our starting point $\vec{\lambda}^0$.
The variance is more tedious. Let's follow Dodelson's exercise and restrict ourselves to one parameter λ.
In addition, assume that we are evaluating in the true maximum $\bar{\lambda}$, i.e that $\lambda^0 = \bar{\lambda}$.
Then:

$$\langle (\hat{\lambda} - \bar{\lambda})^2 \rangle = \langle \hat{\lambda}^2 \rangle - 2 \langle \hat{\lambda} \rangle \bar{\lambda} + \bar{\lambda}^2 = \langle \hat{\lambda}^2 \rangle - \bar{\lambda}^2$$

$$= \Big\langle \Big(\lambda^0 + \tfrac{1}{2} F^{-1}_{\lambda\lambda} \big[\cancel{\mathrm{Tr}}\, \Delta C^{-1} C_{,\lambda} C^{-1} \Delta - \mathrm{Tr}\, C^{-1} C_{,\lambda} \big] \Big)$$
$$\Big(\lambda^0 + \tfrac{1}{2} F^{-1}_{\lambda\lambda} \big[\Delta C^{-1} C_{,\lambda} C^{-1} \Delta - \mathrm{Tr}\, C^{-1} C_{,\lambda} \big] \Big) \Big\rangle - \bar{\lambda}^2$$

$$= (\lambda^0)^2 - \bar{\lambda}^2 + \lambda^0 F^{-1}_{\lambda\lambda} \big[\langle \Delta C^{-1} C_{,\lambda} C^{-1} \Delta \rangle - \mathrm{Tr}\, C^{-1} C_{,\lambda} \big]$$
$$+ \tfrac{1}{4} F^{-2}_{\lambda\lambda} \Big\langle \big[\Delta C^{-1} C_{,\lambda} C^{-1} \Delta - \mathrm{Tr}\, C^{-1} C_{,\lambda} \big]$$
$$\big[\Delta C^{-1} C_{,\lambda} C^{-1} \Delta - \mathrm{Tr}\, C^{-1} C_{,\lambda} \big] \Big\rangle$$

Remember: $\langle \Delta A \Delta \rangle = \mathrm{Tr}\, C A$

in addition, as the likelihood is gaussian:

$$\langle \Delta_i \Delta_j \Delta_k \Delta_l \rangle = C_{ij} C_{kl} + C_{ik} C_{jl} + C_{il} C_{jk}$$

, i.e. the 4 point function is expressed roughly speaking as

$\langle \Delta^4 \rangle = 3 \langle \Delta^2 \rangle \langle \Delta^2 \rangle$

Using $\langle \Delta C^{-1} C_{,\lambda} C^{-1} \Delta \rangle = \mathrm{Tr}\, C^{-1} C_{,\lambda}$ cancels the cross terms, so we are left with

$$\langle (\hat{\lambda} - \bar{\lambda})^2 \rangle = \frac{1}{4} F_{\lambda\lambda}^{-2} \Big[\big\langle \Delta_i (C^{-1} C_{,\lambda} C^{-1})_{ij} \Delta_j \Delta_k (C^{-1} C_{,\lambda} C^{-1})_{kl} \Delta_l \big\rangle$$

$$- 2\, \mathrm{Tr}\, C^{-1} C_{,\lambda}\, \mathrm{Tr}\, C^{-1} C_{,\lambda} + \mathrm{Tr}\, C^{-1} C_{,\lambda}\, C^{-1} C_{,\lambda} \Big]$$

$$= \frac{1}{4} F_{\lambda\lambda}^{-2} \Big[(C^{-1} C_{,\lambda} C^{-1})_{ij} (C^{-1} C_{,\lambda} C^{-1})_{kl} (C_{ij} C_{kl} + C_{ik} C_{jl} + C_{il} C_{jk})$$

$$- \mathrm{Tr}\, C^{-1} C_{,\lambda}\, \mathrm{Tr}\, C^{-1} C_{,\lambda} \Big]$$

The $C_{ij} C_{kl}$ term leads to $\mathrm{Tr}\, C^{-1} C_{,\lambda}\, \mathrm{Tr}\, C^{-1} C_{,\lambda}$ which cancels the last term in [ ] brackets:

$$\langle (\hat{\lambda} - \bar{\lambda})^2 \rangle = \frac{1}{4} F_{\lambda\lambda}^{-2} \Big[(C^{-1} C_{,\lambda} C^{-1})_{ij} (C^{-1} C_{,\lambda} C^{-1})_{kl} (C_{ik} C_{jl} + C_{il} C_{jk}) \Big]$$

Both C as well as $C^{-1} C_{,\lambda} C^{-1}$ are symmetric.
Hence, the term in square bracket is:

$$2 \big[C (C^{-1} C_{,\lambda} C^{-1}) C (C^{-1} C_{,\lambda} C^{-1}) \big]_{kk} \quad \text{(sum)}$$

$$= 2\, \mathrm{Tr}\, C (C^{-1} C_{,\lambda} C^{-1}) C (C^{-1} C_{,\lambda} C^{-1})$$

$$= 2\, \mathrm{Tr}\, C^{-1} C_{,\lambda} C^{-1} C_{,\lambda}$$

all in all, we hence get

$$\langle (\hat{\lambda} - \bar{\lambda})^2 \rangle = \frac{1}{2} F_{\lambda\lambda}^{-2} \, \mathrm{Tr}\, C^{-1} C_{,\lambda} C^{-1} C_{,\lambda}$$

$$= F_{\lambda\lambda}^{-2} F_{\lambda\lambda} = \underline{\underline{F_{\lambda\lambda}^{-1}}}$$

So the error of our parameter estimation is $F_{\lambda\lambda}^{-1/2}$. We can hence compute the errors by computing F. For this, we do not need any data, because F is given by $\mathrm{Tr}\, C^{-1} C_{,\lambda} C^{-1} C_{,\lambda}$. This is very useful for forecasting how well upcoming experiments can constrain some parameters.

(DOD) 11.4. The Fisher matrix: Limits and applications

We will use the Fisher matrix for parameter forecasting. We will, however only consider the CMB. You can look up LSS results in Dodelson's book

11.4.1. CMB Fisher matrix

We need to answer two questions:

1. What pixels do we use?
2. What parameters do we want to measure?

We would like to measure the multipole coefficients C_ℓ, which we will call λ_ℓ to avoid confusion with C. Hence, it's best to use the $a_{\ell m}$'s as pixels, i.e. we use

$$a_{\ell m} = \int d\Omega \, Y_\ell^{m*}(\vec{n}) \, \Theta(\vec{n})$$

and the covariance matrix looks as follows:

$$C_{\ell m \ell' m'} = \langle a_{\ell m} a^*_{\ell' m'} \rangle$$

We have already seen the effect of a window function of beam size σ on the signal covariance matrix. The signal covariance is hence

$$C_S = \delta_{\ell\ell'} \delta_{mm'} e^{-\ell^2 \sigma^2} \lambda_\ell$$

What about the noise? Let's follow Dodelson and assume that our satellite measures pixels of angular extent $\Delta\Omega$ which have gaussian noise σ_n that is uncorrelated between the pixels:

$$\langle \Theta(\vec{n}_i)\,\Theta(\vec{n}_j)\rangle_{noise} = \delta_{ij}\,\sigma_n^2 \qquad \text{for pixels pointing } \vec{n}_i, \vec{n}_j$$

$$C_N = \langle a_{\ell m}\, a^*_{\ell' m'}\rangle = \sum_{i,j} (\Delta\Omega)^2\, Y_\ell^{m*}(\vec{n}_i)\, Y_{\ell'}^{m'}(\vec{n}_j)\, \langle \Theta(\vec{n}_i)\,\Theta(\vec{n}_j)\rangle$$

$$= \sum_i (\Delta\Omega)^2\, Y_\ell^{m*}(\vec{n}_i)\, Y_{\ell'}^{m'}(\vec{n}_j)\, \sigma_n^2$$

moving back to integrals: $\sum_i \Delta\Omega \rightarrow \int d\Omega$

we get

$$C_N = \Delta\Omega\,\sigma_n^2 \underbrace{\int d\Omega\, Y_\ell^{m*}(\vec{n})\, Y_{\ell'}^{m'}(\vec{n})}_{=\delta_{\ell\ell'}\delta_{mm'}\ \text{(normalization)}}$$

$$= \Delta\Omega\,\sigma_n^2\,\delta_{\ell\ell'}\delta_{mm'} \Rightarrow \text{double the pixels} \Rightarrow \Delta\Omega \rightarrow \Delta\Omega/2$$

$$\Rightarrow \sigma_n^2 \rightarrow 2\sigma_n^2 \Rightarrow \text{invariant}$$

So the total covariance is

$$C_{\ell m \ell' m'} = \delta_{\ell\ell'}\,\delta_{mm'}\left[\lambda_\ell\, e^{-\ell^2\sigma^2} + w^{-1}\right]$$

where $w^{-1} = (\sigma_n^2\,\Delta\Omega)$

The inverse is

$$(C^{-1})_{\ell m \ell' m'} = \delta_{\ell\ell'}\,\delta_{mm'}\left[\lambda_\ell\, e^{-\ell^2\sigma^2} + w^{-1}\right]^{-1}$$

and the derivative w.r.t. the parameters λ_α is:

$$C_{\ell m \ell' m', \alpha} = \delta_{\ell\ell'}\delta_{mm'}\delta_{\ell\alpha}\, e^{-\ell^2\sigma^2}$$

So the Fisher matrix is:

$$F_{\alpha\alpha'} = \frac{1}{2}\,\mathrm{Tr}\; C^{-1} C_{,\alpha} C^{-1} C_{,\alpha'}$$

$$= \frac{1}{2}\, C_{\ell m \ell' m', \alpha}\; C^{-1}_{\ell' m' \ell'' m''}\; C_{\ell'' m'' \ell''' m''', \alpha'}\; C^{-1}_{\ell''' m''' \ell m}$$

$$= \frac{1}{2}\,\delta_{\ell\ell'}\delta_{mm'}\delta_{\ell\alpha}\, e^{-\ell^2\sigma^2}\;\frac{\delta_{\ell'\ell''}\,\delta_{m'm''}}{\lambda_{\ell'} e^{-\ell'^2\sigma^2}+w^{-1}}\;\delta_{\ell''\ell'''}\,\delta_{m''m'''}\,\delta_{\ell''\alpha'}\, e^{-\ell''^2\sigma^2}$$

$$\frac{\delta_{\ell'''\ell}\,\delta_{m'''m}}{\lambda_\ell e^{-\ell^2\sigma^2}+w^{-1}} \qquad \text{(sum assumed)}$$

Consider first sum over m, m', m'', m''':

$$\sum_{m m' m'' m'''} \delta_{mm'}\,\delta_{m'm''}\,\delta_{m''}\delta_{m'''}\,\delta_{m'''}\delta_{m} = \delta_{mm} = \underline{\underline{2\ell+1}}$$

Then sum over ℓ', ℓ'', ℓ''':

$$\sum_{\ell\ell'\ell''\ell'''} \delta_{\ell\ell'}\,\delta_{\ell\alpha}\,\delta_{\ell'\ell''}\,\delta_{\ell''\ell'''}\,\delta_{\ell''\alpha'}\,\delta_{\ell'''\ell}$$

$$= \delta_{\alpha\alpha'} \qquad (\text{and } \alpha=\ell,\ \alpha'=\ell)$$

So in the all-sky limit, the Fisher matrix becomes:

$$F_{\ell\ell'} = \frac{2\ell+1}{2}\,\delta_{\ell\ell'}\, e^{-2\ell^2\sigma^2}\left[C_\ell e^{-\ell^2\sigma^2}+w^{-1}\right]^{-2}$$

substituted $\lambda_\ell \to C_\ell$ name back again

So in an all sky survey, there is no correlation between different ℓ's and the error on determining C_ℓ is given by $\sqrt{F^{-1}}$, so:

$$\delta C_\ell = \sqrt{\frac{2}{2\ell+1}} \left[C_\ell + w^{-1} e^{\ell^2 \sigma^2} \right]$$

Where the $\sqrt{\frac{2}{2\ell+1}}\, C_\ell$ is the cosmic variance uncertainty and the second term the effect of noise and beam width. The factor $(2\ell+1)$ which beats down the error for large ℓ from cosmic variance is there, because for any ℓ, we have $(2\ell+1)$ independent $a_{\ell m}$'s to estimate it.

As no experiment can cover the full sky (foregrounds etc), the error for an experiment of sky fraction f_{sky} is:

$$\delta C_\ell = \sqrt{\frac{2}{(2\ell+1) f_{sky}}} \left[C_\ell + w^{-1} e^{\ell^2 \sigma^2} \right]$$

So what matters here are: f_{sky}, beam width σ and weight w.