CMB Map making

Paper: M. Tegmark, astro-ph/9611130

CMB experiments scan the sky in some region for some time (from days to years - balloons / satellite) to gather intensity of microwave radiation.

Time ordered data can be represented as:

$$d_t = P_{ti}\, s_i + n_t$$

sky signal of pixel i

noise

time ord. data

Pointing matrix: "Where does satellite point to at time t?"

$$\vec{d} = P\vec{s} + \vec{n} \qquad \text{in matrix notation}$$

Noise is supposed to ~~have~~ zero mean

$\langle \vec{n} \rangle = 0$ and covariance

$\langle \vec{n}\vec{n}^T \rangle \equiv N$ known

As we will see, many methods of Map making available. We will discuss COBE method. We find it by minimizing the χ^2

$$\chi^2 = (\vec{d} - P\vec{s})^T N^{-1} (\vec{d} - P\vec{s})$$

$$= \sum_{i,j,t,t'} (d_t - P_{ti} s_i)(N^{-1})_{tt'} (d_{t'} - P_{t'j} s_j)$$

$$\frac{\partial \chi^2}{\partial s_i} = -2 \sum_{tt'j} P_{ti} N^{-1}_{tt'} (d_{t'} - P_{t'j} s_j) \stackrel{!}{=} 0$$

for minimal χ^2. So

$$\sum_{tt'j} P_{ti} N^{-1}_{tt'} d_{t'} = \sum_{tt'j} P_{ti} N^{-1}_{tt'} P_{t'j} s_j$$

$$P^T N^{-1} \vec{d} = \underbrace{P^T N^{-1} P}_{\equiv C_N^{-1}} \vec{s} \qquad \text{call this } C_N \cdots$$

multiply by C_N, then

$$C_N P^T N^{-1} \vec{d} = \vec{s}$$

is the minimal χ^2 value, which we will call Δ, so

$$\vec{\Delta} = C_N P^T N^{-1} \vec{d} = [P^T N^{-1} P]^{-1} P^T N^{-1} \vec{d}$$

This is a <u>linear</u> method, because our estimate of the temperature anisotropy at some pixel is a linear function of $\vec{d}$. Following Tegmark, let's denote this as

$$\vec{\Delta} = W \vec{d}$$

and in our Cobe case, $W = [P^T N^{-1} P]^{-1} P^T N^{-1}$

but there are many more choices possible for W.

No.	Method	Specification
1	Generalized COBE	$\mathbf{W} = [\mathbf{A}^t\mathbf{M}\mathbf{A}]^{-1}\mathbf{A}^t\mathbf{M}$
2	Bin averaging	$\mathbf{W} = [\mathbf{A}^t\mathbf{A}]^{-1}\mathbf{A}^t$
3	COBE	$\mathbf{W} = [\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]^{-1}\mathbf{A}^t\mathbf{N}^{-1}$
4	Wiener 1	$\mathbf{W} = \mathbf{S}\mathbf{A}^t[\mathbf{A}\mathbf{S}\mathbf{A}^t + \mathbf{N}]^{-1}$
5	Wiener 2	$\mathbf{W} = [\mathbf{S}^{-1} + \mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]^{-1}\mathbf{A}^t\mathbf{N}^{-1}$
6	Saskatoon	$\mathbf{W} = [\eta\mathbf{S}^{-1} + \mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]^{-1}\mathbf{A}^t\mathbf{N}^{-1}$
7	TE96	$\mathbf{W} = \Lambda\mathbf{S}\mathbf{A}^t[\mathbf{A}\mathbf{S}\mathbf{A}^t + \mathbf{N}]^{-1}, \quad (\mathbf{W}\mathbf{A})_{ii} = 1$
8	TE97	$\mathbf{W} = \Lambda[\eta\mathbf{S}^{-1} + \mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]^{-1}\mathbf{A}^t\mathbf{N}^{-1}, \quad (\mathbf{W}\mathbf{A})_{ii} = 1$
9	Maximum probability	Nonlinear method if non-Gaussian
10	Maximum entropy	Nonlinear method

Table 1: Map-making methods

be augmented to include the brightness of various foreground components in each pixel, and the matrix $\mathbf{A}$ would encompass the assumptions made about their frequency dependence.

Without loss of generality, we can take the noise vector to have zero mean, i.e., $\langle\mathbf{n}\rangle = 0$, so the noise covariance matrix is

$$\mathbf{N} \equiv \langle\mathbf{n}\mathbf{n}^t\rangle. \tag{2}$$

In some of the methods described below (methods 4-9), the following prior assumptions are made about the map: it is assumed to be a realization of random vector with zero mean, i.e., $\langle\mathbf{x}\rangle = 0$, with some known covariance matrix

$$\mathbf{S} \equiv \langle\mathbf{x}\mathbf{x}^t\rangle \tag{3}$$

and uncorrelated with the noise, i.e., $\langle\mathbf{n}\mathbf{x}^t\rangle = 0$.

2.2 Ten mapping methods

We will now summarize some map-making methods that have recently been used or advocated in the CMB context. All *linear* methods can clearly be written in the form

$$\tilde{\mathbf{x}} = \mathbf{W}\mathbf{y}, \tag{4}$$

where $\tilde{\mathbf{x}}$ denotes the estimate of the map $\mathbf{x}$ and $\mathbf{W}$ is some $m \times n$ matrix that specifies the method. Table 1 shows the choices of $\mathbf{W}$ that define the linear methods we will discuss.

Let us compute the error of our estimate Δ for s:

$$\vec{\varepsilon} \equiv \vec{\Delta} - \vec{s} = W\vec{d} - \vec{s} = W[P\vec{s} + \vec{n}] - \vec{s}$$

$$= [WP - \mathbb{1}]\vec{s} + W\vec{n}$$

but for methods for which $WP = \mathbb{1}$, like in our case:

$$WP = \underbrace{[P^T N^{-1} P]^{-1}\, P^T N^{-1} P}_{= \mathbb{1}}$$

this error is independent of the signal

$$\vec{\varepsilon} = W\vec{n},$$

i.e. $\vec{\Delta}$ is $\vec{s}$ plus some noise independent of signal.

Noise covariance matrix

$\langle \vec{\varepsilon}\vec{\varepsilon}^{\,T} \rangle$ is indeed $[P^T N^{-1} P]^{-1} = C_N$

So subscript N of C_N is justified ☺ Proof:

$$\langle \vec{\varepsilon}\vec{\varepsilon}^{\,T} \rangle = \langle W\vec{n}(W\vec{n})^T \rangle = \langle W\vec{n}\vec{n}^T W^T \rangle$$

$$= \left\langle [P^T N^{-1} P]^{-1} P^T \underbrace{N^{-1}\vec{n}\vec{n}^T}_{=\mathbb{1}} (N^{-1})^T P \,[P^T N^{-1} P]^{-1\,T} \right\rangle$$

$$= [P^T N^{-1} P]^{-1} P^T (N^{-1})^T P\, [P^T N^{-1} P]^{-1\,T}$$

use $(N^{-1})^T = N^{-1}$ for any symmetric matrix, proof soon

and $P^T N^{-1} P$ symmetric (below),

then:

$$\langle \vec{\varepsilon}\varepsilon^T \rangle = [P^T N^{-1} P]^{-1} \underbrace{P^T N^{-1} P\, [P^T N^{-1} P]^{-1}}_{= \mathbb{1}}$$

$$= [P^T N^{-1} P]^{-1}$$

$$= C_N \quad \checkmark$$

Proof for $A^T = A \Rightarrow (A^{-1})^T = A^{-1}$:

$$\mathbb{1} = \mathbb{1}^T = A^{-1} A = \left(A^T (A^{-1})^T\right)^T = \mathbb{1}$$ transpose again:

$$\mathbb{1}^T = A^T (A^{-1})^T = A (A^{-1})^T \Rightarrow \underline{\underline{(A^{-1})^T = A^{-1}}}$$

Proof for $P_T N^{-1} P$ symmetric:

Take any matrix $A = A^T$, then

$$B^T A B = B^T A^T B = (B^T A B)^T \quad \text{q.e.d.}$$

Remark for useful identity we need soon:

$$[A + B]^{-1} = [\mathbb{1} + A^{-1} B]^{-1} A^{-1} \text{, because}$$

$$[A+B][A+B]^{-1} = [A+B][\mathbb{1} + A^{-1} B]^{-1} A^{-1}$$

$$= A \underbrace{[\mathbb{1} + A^{-1} B][\mathbb{1} + A^{-1} B]^{-1}}_{\mathbb{1}} A^{-1} = A A^{-1} = \mathbb{1} \text{ ; q.e.d.}$$

Quick summary, as lecture 2 weeks ago:

We make temperature anisotropy map for CMB with pixel data s_i of Pixel i given time stream of data

$\vec{d} = P\vec{s} + \vec{n}$; typically $\vec{d}$ has $\sim 10^8$ to 10^{10} entries
$\vec{s}$ has $\sim 10^3$ to 10^6 entries

timeordered data, pointing, noise

$\langle \vec{n}\vec{n}^T \rangle = N$ diagonal noise covariance given

We showed that COBE method yields estimate of $\vec{s}$, called $\vec{\Delta}$ which minimizes

$$\chi^2 = (\vec{d} - P\vec{s})^T N^{-1} (\vec{d} - P\vec{s})$$

namely
$$\vec{\Delta} = C_N P^T N^{-1} \vec{d} \equiv [P^T N^{-1} P]^{-1} P^T N^{-1} \vec{d}$$
$$\equiv W\vec{d}$$

Show Tegmark slide!

We proofed that $\vec{\varepsilon} \equiv \vec{\Delta} - \vec{s}$ has covariance

$$\langle \vec{\varepsilon}\vec{\varepsilon}^T \rangle = [P^T N^{-1} P]^{-1} = C_N$$

The Fisher Matrix – first encounter

Rather unpedagogically, let me state that the Fisher matrix F is given by

$$F_{ij} \equiv -\left\langle \frac{\partial^2 \ln \mathrm{prob}(\vec{\lambda}|\vec{x})}{\partial\lambda_i\,\partial\lambda_j} \right\rangle$$

$$= -\langle L^{(2)} \rangle$$

Where $\vec{\lambda}$ are the parameters and $\vec{x}$ is the data.
It is also called the Fisher information matrix.
There is a theorem, called "Cramer-Rao" inequality which says that no method can measure a parameter i better than the square root of the diagonal element of F^{-1}, i.e

$$\Delta\lambda_i \geqslant \sqrt{(F^{-1})_{ii}}$$

In case someone tells you λ_j, $j \neq i$, i.e. you have perfect prior knowledge of λ_j except $j=i$, the error is somewhat smaller

$$\Delta\lambda_i \geqslant \sqrt{(F_{ii})^{-1}}$$

For a Gaussian distribution with zero mean and covariance matrix C, we will later derive that

$$F_{ij} = \frac{1}{2} \mathrm{tr}\, C^{-1} C_{,i}\, C^{-1} C_{,j} \qquad \frac{dC}{d\lambda_i} = C_{,i}$$

A lossles map

We will show that making the map from TOD, did not destroy any information about cosmological parameters. Let us follow Tegmark and call

$$G_i \equiv C^{-1} C_{,i}$$

As a first observation, take a look at ~~an~~ a ~~orthogonal~~ transformation of $\vec{\Delta} = W\vec{d}$:

$$\vec{\Delta}' = B\vec{\Delta} \qquad B \text{ invertible}$$

The temperature anisotropies are themselves gaussian distributed. This is a "prediction" of inflation, but must always be tested of course by our CMB experiments. So far, no-one has found any deviation from

$$\mathrm{prob}(\Delta) \propto \exp\left(-\frac{1}{2} \vec{\Delta}^t C^{-1} \Delta\right)$$

which we will derive soon.

So $$e^{-\vec{\Delta}'^T C'^{-1} \vec{\Delta}'} = e^{-\vec{\Delta}^T B^T C'^{-1} B \vec{\Delta}} \overset{!}{=} e^{-\vec{\Delta}^T C^{-1} \vec{\Delta}}$$

$$\Rightarrow B^T C'^{-1} B \overset{!}{=} C^{-1} \Rightarrow C'^{-1} = B^{T^{-1}} C^{-1} B$$

for $\Rightarrow C' = B C B^T$ and hence

$$G'_i = C'^{-1} C'_{,i} = (B^T)^{-1} C^{-1} \cancel{B^{-1} B} C_{,i} B^T$$
$$= (B^T)^{-1} C^{-1} C_{,i} B^T$$

But then

$$F'_{ij} = \frac{1}{2} \operatorname{tr} G'_i G'_j = \frac{1}{2} \operatorname{tr} (B^T)^{-1} C^{-1} C_{,i} B^T (B^T)^{-1} C^{-1} C_{,j} B^T$$

but trace is cyclic

$$\Rightarrow F'_{ij} = \frac{1}{2} \operatorname{tr} \cancel{B^T (B^T)^{-1}} C^{-1} C_{,i} \cancel{B^T (B^T)^{-1}} C^{-1} C_{,j}$$
$$= \frac{1}{2} \operatorname{tr} C^{-1} C_{,i} C^{-1} C_{,j} = F_{ij}$$

So methods 3-8 are information-theoretically equivalent, because these matrices W which give estimates $\vec{\Delta} = W \vec{d}$ are all equivalent by multiplying a suitable B from the left.

Show Table of Tegmark.

Methods 3-8 lose no information

Cobe Method, i.e. method 3 was

$$W = [P^T N^{-1} P]^{-1} P^T N^{-1}$$

As multiplying by B leaves result invariant, We can simplify by multiplying $\Sigma^{-1} \equiv [P^T N^{-1} P]$ from left and use

$$W = P^T N^{-1}$$

$$C^{map} = \langle \vec{\Delta} \Delta^T \rangle = \qquad \langle W\vec{d}\,(W\vec{d})^T \rangle$$

$$= \langle (W[P\vec{s}+\vec{n}]\,(W[P\vec{s}+\vec{n}])^T \rangle$$

$$= \langle (WP\vec{s} + W\vec{n})(\vec{n}^T W^T + \vec{s}^T P^T W^T) \rangle$$

$$= \langle WP\vec{s}\vec{s}^T P^T W^T + \qquad WP\vec{s}\vec{n}^T W^T$$

$$+ W\vec{n}\vec{n}^T W^T + W\vec{n}\vec{s}^T P^T W^T \rangle \quad ; \langle \vec{n}\vec{s}^T \rangle = \langle \vec{s}\vec{n}^T \rangle = 0$$

$$= WP \underbrace{\langle s\vec{s}^T \rangle}_{S} P^T W^T + W \underbrace{\langle \vec{n}\vec{n}^T \rangle}_{N} W^T$$

$$= P^T N^{-1} P S P^T \underbrace{(N^{-1})^T}_{N^{-1}} P + P^T \cancel{N^{-1} N} \underbrace{(N^{-1})^T}_{N^{-1}} P$$

$$= P^T N^{-1} P S P^T N^{-1} P + P^T N^{-1} P$$

$$= \Sigma^{-1} [\mathbb{1} + S \Sigma^{-1}]$$

$$C^{map}_{1i} = \Sigma^{-1} S_{1i} \Sigma^{-1}$$

$$C^{-1}_{map} = [\mathbb{1} + S\Sigma^{-1}]^{-1} \Sigma$$

$$G^{map}_i = C^{-1}_{map} C^{map}_{1i} = [\mathbb{1} + S\Sigma^{-1}]^{-1} \Sigma \Sigma^{-1} S_{1i} \Sigma^{-1}$$

$$= [\mathbb{1} + S\Sigma^{-1}]^{-1} \quad S_{1i} \Sigma^{-1}$$

9 $$= [\mathbb{1} + S\Sigma^{-1}]^{-1} \quad S_{1i} P^T N^{-1} P$$

$$C^{TOD} = \langle \vec{d}\vec{d}^T \rangle = \langle (P\vec{s}+\vec{n})(P\vec{s}+\vec{n})^T \rangle$$

$$= PSP^T + N$$

$$C^{TOD}_{,i} = P S_{,i} P^T$$

$$C^{-1}_{TOD} = [PSP^T+N]^{-1}$$

" $\frac{1}{N+PSP^T} = \frac{1}{N[1+N^{-1}PSP^T]}$ "

$$G^{TOD}_i = [PSP^T+N]^{-1} P S_{,i} P^T$$

$$= [\mathbb{1} + N^{-1}PSP^T]^{-1} N^{-1} P S_{,i} P^T$$

Use geometric series $[\mathbb{1}+M]^{-1} = 1 - M + M^2 - M^3 + \ldots$

$$\Rightarrow G^{TOD}_i = \left[\mathbb{1} - N^{-1}PSP^T + N^{-1}PSP^T N^{-1}PSP^T - \ldots\right] N^{-1}PS_{,i}P^T$$

$$= N^{-1}P\left[\mathbb{1} - SP^TN^{-1}P + SP^TN^{-1}PSP^TN^{-1}P - \ldots\right] S_{,i}P^T$$

$$= N^{-1}P[\mathbb{1} + SP^TN^{-1}P]^{-1} S_{,i}P^T$$

$$= N^{-1}P[\mathbb{1} + S\Sigma^{-1}]^{-1} S_{,i} P_T$$

$$F_{ij}^{map} = \frac{1}{2}\mathrm{tr}\, [\mathbb{1}+S\Sigma^{-1}]^{-1} S_{,i}\, P^TN^{-1}P\, [\mathbb{1}+S\Sigma^{-1}]^{-1} S_{,j} P^TN^{-1}P$$

$$F_{ij}^{tod} = \frac{1}{2}\mathrm{tr}\, \cancel{\Sigma}\, N^{-1}P[\mathbb{1}+S\Sigma^{-1}]^{-1} S_{,i}\, P^TN^{-1}P\, [\mathbb{1}+S\Sigma^{-1}]^{-1} S_{,j} P^T$$

$\Rightarrow$ trace cyclic $\Rightarrow$ $\underline{\underline{F^{tod} = F^{map}}}$?

So methods 3–8 ~~loose~~ lose no information?