Making maps for CMB experiments

Reijo Keskitalo[*]

LBL, US

E-mail: RTKeskitalo@lbl.gov

Elina Keihänen

University of Helsinki, Finland

Hannu Kurki-Suonio

University of Helsinki, Finland

Torsti Poutanen

University of Helsinki, Finland

Anna-Stiina Suur-Uski

University of Helsinki, Finland

Julian Borrill

LBL, US

Theodore Kisner

LBL, US

We review the mapmaking problem in the context of CMB experiments and present the spectrum of different techniques available. We then draw analogs between these methods and elementary signal processing concepts and explain where such analogies work and where they do not.

---

[*]Speaker.

1. Background

Mapmaking in the cosmic microwave background (CMB) context refers to the process of reducing a set of instrumental observations of the microwave sky into maps or other sky representations. The sky representations are further processed to extract information about the structure of anisotropies present in the microwave background. CMB mapmaking distinguishes itself from other astrophysical imaging by having

- a wide field of view (from tens of square degrees to full sky)

- wide instrumental beam (from few to several tens of arc minutes)

- wide bandpass (order of 10 % of the center frequency)

- small number of detectors or pixels (from a few to a few thousand)

- scientific signal at best 10 % of the foreground anisotropy (kinematic dipole vs. primary anisotropy) and only $1/100\,000$ of the background (monopole vs. primary anisotropy)

- significant noise autocorrelations

Because of the above characteristics, mapmakers are very concerned about extracting the maximal information content out of the observations. Both beam and bandpass effects would require their own treatments to be exhaustive so in this text we will focus on the optimal treatment of correlated instrumental noise.

We recognize two main classes of instrumental noise correlations: low and high frequency. The division is somewhat arbitrary but one can consider the sky signal to manifest in the detector streams at a certain frequency band, say between 10 mHz and 10 Hz. Noise power below this band is "low frequency" and noise above this frequency is "high frequency". These numbers are just orders of magnitudes and are affected by the scanning frequency and the width of the instrumental beam. Low frequency noise correlations are driven by the electronic amplification of the signal and usually fit well the $1/f$ noise model:

$$P(f) \propto \left(\frac{f}{f_{\mathrm{k}}}\right)^{\alpha}, \tag{1.1}$$

where the *noise power spectral density*, P, is a function of frequency, knee frequency and slope (f, f_{k} and α respectively). Barring any harmonics from the electronics and cooling systems, the instrumental noise spectrum at high frequency is close to flat (white). For most radiometer detectors this means that the overall noise spectrum is well characterized by

$$P(f) = \frac{\sigma^2}{f_{\mathrm{s}}}\left[1 + \left(\frac{f}{f_{\mathrm{k}}}\right)^{\alpha}\right], \tag{1.2}$$

where σ is the *white noise* standard deviation and f_{s} is the sampling frequency.

Some detector technologies introduce a lag in the detector response. In bolometers, incident radiation heats up the target and depending of the amount of absorbed intensity, it takes a while for the heat to dissipate. It is possible to measure and correct for such effects but it comes at a cost.

Because of the bolometer response deconvolution, bolometer noise spectra can have high frequency features that do not fit the simple $1/f$ noise model. Examples of radiometer and bolometer noise power spectral densities are given in Figure 1.
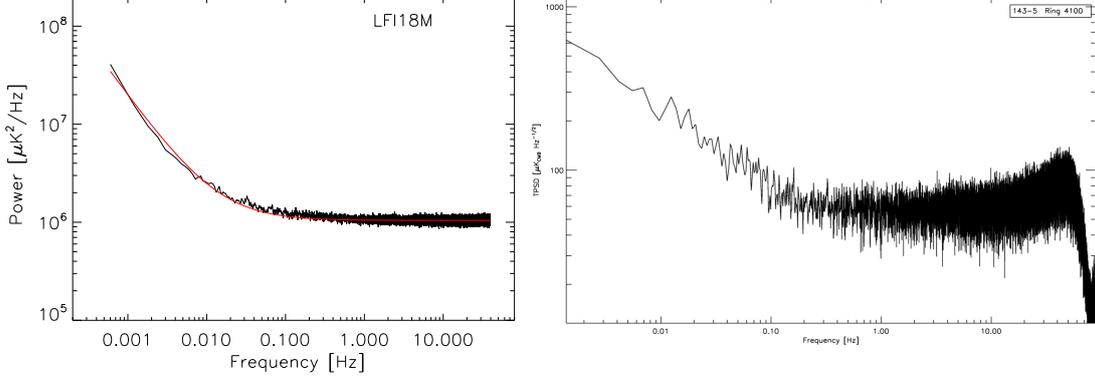
Figure 1: Example radiometer (left) and bolometer (right) noise spectra from [1] and [2] respectively.

2. Primer

The goal of mapmaking is to evaluate an estimator of the sky representation, $\widetilde{\mathbf{m}}$, from a vector of observations, $\mathbf{y}$. Typically, the sky representation, $\mathbf{m}$, is taken to be a vector of pixel values but pixelized maps are not a requirement. For full sky analysis of statistically isotropic CMB, coefficients of the spherical harmonic expansion of the sky are often more useful than pixel values.

Relationship between the observations, $\mathbf{y}$, and the sky is expressed as

$$\mathbf{y} = \mathbf{Pm} + \mathbf{n}, \tag{2.1}$$

where $\mathbf{P}$ is the *pointing matrix*. It maps the sky representation into individual detector samples. The additional term, $\mathbf{n}$, is a vector of instrumental noise.

There is a subtle interplay between the pointing matrix and the sky representation. Physically each detector sample is a 4π convolution of the instrumental beam response over the sky and a frequency band convolution of the detector bandpass and the frequency dependent sky. The beam response causes an astrophysical object like the Crab nebula, M1, that has a beautiful, complex structure in the Hubble Space Telescope images, appear as a deformed blob in CMB maps. This poses no problem as the physical processes behind CMB fluctuations ensure absence of structure at these scales.

The bandpass affects the apparent intensity of a source with non-flat spectrum. For example the radio galaxy 3C 405 appears as one of the brightest compact sources on the sky through a 30GHz detector but is just detectable using an 857GHz detector. Even detectors that are designed and built to have similar bandpass end up measuring a slightly different sky emission due to the effective bandpass being slightly different.

Although it is possible to construct $\mathbf{P}$ to include both beam and bandpass, solving for the fiducial sky would become intractable. It is customary to simplify the problem by assigning parts of

the instrumental response into the sky representation. For example, modeling the beam response as a δ-function in **P** yields a smoothed sky representation that is convolved with the effective instrumental beam. For most analysis this is entirely acceptable as long as the effect is well quantified. The bandpass effects are equally disregarded in most mapmaking by only considering detectors that have similar bandpasses.

In what follows we will take **P** to be a simple pixelized pointing matrix where the signal part of a sample is just the corresponding pixel value. In case of polarized detectors, the signal is a linear combination of the intensity and polarization pixel values.

Provided that the noise part, **n**, can be modeled as Gaussian (typically an excellent approximation), one can solve for a maximum likelihood estimator [6] for **m**:

$$\widetilde{\mathbf{m}} = \left(\mathbf{P}^{\mathrm{T}}\mathbf{N}^{-1}\mathbf{P}\right)^{-1}\mathbf{P}^{\mathrm{T}}\mathbf{N}^{-1}\mathbf{y}, \tag{2.2}$$

where we have written $\mathbf{N} = \langle \mathbf{n}\mathbf{n}^{\mathrm{T}} \rangle$ for the zero mean noise covariance matrix.

In practice, explicitly evaluating Eq. (2.2) is often impossible due to the large dimensions of **P** and **N**. Instead, $\widetilde{\mathbf{m}}$ is gradually refined through conjugate gradient iteration until the residual in

$$\mathbf{P}^{\mathrm{T}}\mathbf{N}^{-1}\mathbf{P}\widetilde{\mathbf{m}} = \mathbf{P}^{\mathrm{T}}\mathbf{N}^{-1}\mathbf{y} \tag{2.3}$$

is reduced to acceptable level. The virtue of Eq. (2.3) is that the large matrices need not be evaluated explicitly but it suffices to have them act on vectors. Furthermore, assuming noise to be piece-wise stationary and the stationary intervals uncorrelated allows multiplications by $\mathbf{N}^{-1}$ to be carried out in $\mathcal{O}(n\log n)$ time by performing them in Fourier domain. The δ-function approximation to pointing makes applications of **P** linear.

3. The Spectrum of Methods

The maximum likelihood equation, Eq. (2.2) defines a linear mapmaking operator,

$$\mathbf{M}(\mathbf{P},\mathbf{N}) = \left(\mathbf{P}^{\mathrm{T}}\mathbf{N}^{-1}\mathbf{P}\right)^{-1}\mathbf{P}^{\mathrm{T}}\mathbf{N}^{-1} \tag{3.1}$$

that transforms noisy observations into an estimate of the sky. Most approaches to mapmaking can be viewed as variations of this operator. We have made this explicit by writing **M** to be a function of the pointing matrix, **P**, and the noise model, **N**.

When noise is sufficiently white, it can be approximated by a diagonal covariance matrix, $\mathbf{N} \approx \mathbf{N}_{\mathrm{w}}$. This greatly simplifies the evaluation of the maximum likelihood map, effectively removing the need to iterate the map estimate. An important variation of this simple binning is a case where noise can be high-pass filtered to meet the approximation. Application of such filter will have an impact on signal and it is important to characterize and propagate these effects, typically through Monte Carlo simulations that are very efficient to perform when the mapmaking operator has diagonal noise covariance.

When deviations from white noise spectrum occur only at low frequency, they translate into a slowly varying noise offset in the time streams. If we model the slowly varying component as a step function, we can solve for the step offset values (baselines) much like the pixels in our sky representation. In this *destriping* approach [3], noise is still white, $\mathbf{N} \approx \mathbf{N}_{\mathrm{w}}$, but we have modified

the pointing matrix to include scanning of these offset values: $\mathbf{Pm}$ is replaced by $\mathbf{Pm} + \mathbf{Fa}$ in the signal model. The additional signal-like component, $\mathbf{Fa}$ is composed of the step amplitudes, $\mathbf{a}$, and a baseline pointing matrix, $\mathbf{F}$, that scans the amplitudes into a full time domain vector. It is possible to extend the maximum likelihood formalism by defining an extended pointing matrix, $\mathbf{P}^* := [\mathbf{P}, \mathbf{F}]$ leading to the same form as in Eq. (2.3) but it is more convenient to solve and remove the baseline offsets from the observations and treat the rest of the problem as having just white noise.

Treating the baseline offsets like signal enables us to write the mapmaking equation without any prior knowledge of the baseline distribution. However, if noise estimates are available, it is possible to derive a prior baseline distribution in the form of a baseline-baseline covariance matrix:

$$\mathbf{C}_a := \langle \mathbf{aa}^\mathrm{T} \rangle. \tag{3.2}$$

Now the general noise covariance matrix is not approximated as diagonal. Rather, we prune out the degrees of freedom from the matrix that are not needed to describe the noise:

$$\mathbf{N} \approx \mathbf{F}\,\mathbf{C}_a\,\mathbf{F}^\mathrm{T} + \mathbf{N}_\mathrm{w}. \tag{3.3}$$

Again, it is possible to solve for the maximum likelihood map directly from Eq. (2.3) but it is often simpler to solve for the baseline vector, $\mathbf{a}$, subtract the baselines and use trivial binning to map the signal and whitened noise. It has become customary to call destriping with prior noise information *generalized destriping* [4].

The varying levels of approximations define a spectrum of mapmaking algorithms:

	Binning	Destriping	Generalized destriping	General Least Squares
Signal	$\mathbf{Pm}$	$\mathbf{Pm} + \mathbf{Fa}$	$\mathbf{Pm}$	$\mathbf{Pm}$
Noise	$\mathbf{N}_\mathrm{w}$	$\mathbf{N}_\mathrm{w}$	$\mathbf{F}\,\mathbf{C}_a\,\mathbf{F}^\mathrm{T} + \mathbf{N}_\mathrm{w}$	$\mathbf{N}$

Table 1: Mapmaking approximations

4. Signal Processing Analogy

Destriping, the process of fitting and subtracting baseline offsets, can be considered high pass filtering of the data. When applied to frequencies without signal content (or to a noise only simulation) the analogy works very well. In this case, standard destriping is equivalent to subtracting a running average from the data. In the Fourier domain this corresponds to convolving the signal with a filter:

$$H(f) = \left[1 - \frac{\sin^2(\pi f t_\mathrm{base})}{(\pi f t_\mathrm{base})^2} \right]^2, \tag{4.1}$$

where t_base is the step (baseline) length in seconds. When signal is present, the destriping formulation attempts to apply the above filter only to the noise component of the data. The solution then relies on the constant nature of the sky signal: the part of the observations that varies between samples of the same pixel is noise. There are obvious problems associated with this differentiation: some of the sky signals do change over time, even quite rapidly. The most most prominent examples are distant quasars that can double in intensity over the course of just days. Depending

on the pixel and beam size, samples assigned to the same pixels can also contain traces of signal gradients. For best results, it is customary to exclude from the baseline solution regions of the sky where above problems are likely to interfere.

Another aspect of the above filtering is that it does not differentiate between correlated noise and white noise. If applied to instrument noise spectrum that is predominantly flat (white), the filter still suppresses power below the baseline frequency,

$$f_{\text{base}} = \frac{1}{t_{\text{base}}}. \tag{4.2}$$

Filtering white noise is somewhat of a subtle matter. While it obviously suppresses power in the Fourier domain, it makes the noise have correlations. When the filtered noise is projected onto a map the usual expectation of samples averaging to zero is inhibited. The details depend on the scanning strategy and it is difficult to give general results beyond the correlation argument. Nevertheless, destriping white noise is an example of a situation where the high pass filtering analogy breaks down.

Generalized destriping applies a different kind of filter to the data. It explicitly differentiates between white and correlated noise. Asymptotically (at the limit of short baselines and infinite observations), the generalized filter is

$$H(f) = \frac{P_{\text{w}}^2}{[P_{\text{w}} + P_{\text{c}}(f)]^2}, \tag{4.3}$$

where it is assumed that the total noise spectrum is a sum of the white and correlated noise components, $P(f) = P_{\text{w}} + P_{\text{c}}(f)$. The effects of the destriping filters are depicted in Figure 2.
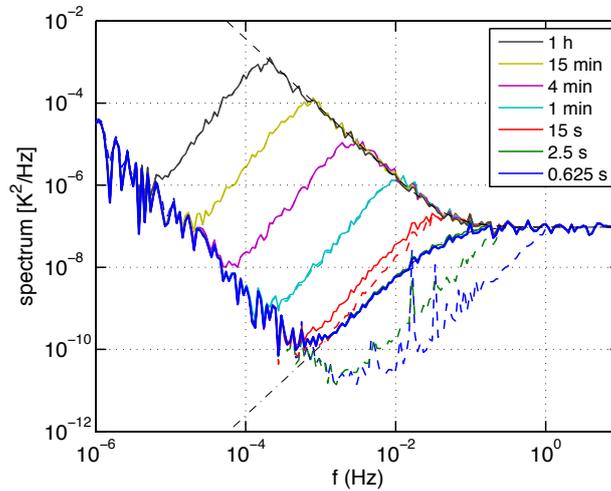
Figure 2: High-pass filtering effect from destriping (dashed lines) and generalized destriping (solid lines). The untreated noise power spectrum is the black dashed line and for baselines longer than 15 seconds the two destriping approaches yield identical results. The image is reproduced from [4].

It turns out that if the step function approximation of the correlated noise is good, the generalized destriping results approach optimal GLS results[5]. This is the case for all instruments with radiometer type noise spectrum. When high frequency correlations are present, analysis has two

paths: leave the high frequency correlations untreated and simply characterize the residual (e.q. through simulations) or apply the optimal GLS method.

References

[1] A. Zacchei et al.: *Planck early results. V. The Low Frequency Instrument data processing*, 2011, A&A **536**, A5 (arXiv:1101.2040)

[2] Planck HFI Core Team: *Planck early results. VI. The High Frequency Instrument data processing*, 2011, A&A **536**, A6 (arXiv:1101.2048)

[3] H. Kurki-Suonio, E. Keihänen, R. Keskitalo, T. Poutanen, D. Maino and C. Burigana: *Destriping CMB temperature and polarization maps*, 2009, A&A, **506**, 1511-1539 (arXiv:0904.3623)

[4] E. Keihänen, R. Keskitalo, H. Kurki-Suonio, T. Poutanen, A.-S. Sirviö: *Making CMB temperature and polarization maps with MADAM*, 2010, A&A, **510**, A57 (arXiv:0907.0367)

[5] M.A.J. Ashdown et al: *Making Maps from Planck LFI 30GHz Data with Asymmetric Beams and Cooler Noise*, 2009, A&A, **493**, 753, (arXiv:0806.3167)

[6] C.M. Cantalupo, J.D. Borrill, A.H. Jaffe, T.S. Kisner, R. Stompor: *MADmap: A Massively Parallel Maximum-Likelihood Cosmic Microwave Background Map-Maker*, 2010, ApJS, **187**, 212, (arXiv:0906.1775)