Map-making problem for CMB data analysis

Thibault Cimic

Laboratoire INRIA - Equipe ALPINES

September 16, 2019

What is CMB ?

CMB : Cosmic Microwave Background

CMB : Cosmic Microwave Background

Other names ?

CMB : Cosmic Microwave Background

Other names ?

Relic radiation

CMB : Cosmic Microwave Background

Other names ?

Relic radiation $\rightarrow$ First photons that started to travel in the very early hot and dense universe (379,000 years old out of 13.8 billions)

CMB : Cosmic Microwave Background
Cosmic : come from far away, outside our galaxy

Other names ?

Relic radiation $\rightarrow$ First photons that started to travel in the very early hot and dense universe (379,000 years old out of 13.8 billions)

CMB : Cosmic Microwave Background
Cosmic : come from far away, outside our galaxy
Microwave : photons as observed today, they lost energy therefore increase wavelengths

Other names ?

Relic radiation $\rightarrow$ First photons that started to travel in the very early hot and dense universe (379,000 years old out of 13.8 billions)

CMB : Cosmic Microwave Background
Cosmic : come from far away, outside our galaxy
Microwave : photons as observed today, they lost energy therefore increase wavelengths
Background : emitted the same way in every direction from anywhere, nearly isotropic and uniform

Other names ?

Relic radiation $\rightarrow$ First photons that started to travel in the very early hot and dense universe (379,000 years old out of 13.8 billions)

CMB : Cosmic Microwave Background
Cosmic : come from far away, outside our galaxy
Microwave : photons as observed today, they lost energy therefore increase wavelengths
Background : emitted the same way in every direction from anywhere, nearly isotropic and uniform

Other names ?

Relic radiation $\rightarrow$ First photons that started to travel in the very early hot and dense universe (379,000 years old out of 13.8 billions)
The goal : Reconstruct a 2D image of "the birth of the Universe" (when it has 0.004% of the current age), i.e. a map of temperature and polarisation of these early photons

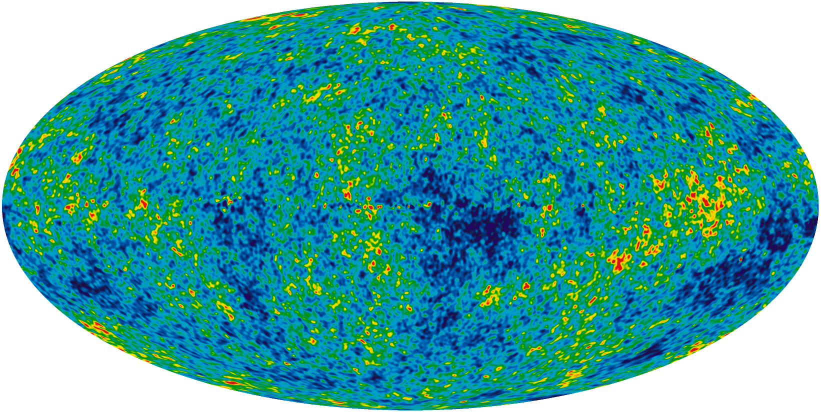


Figure: Map of temperature reconstructed from nine years of WMAP data satellite (2003-2012)

Why is it hard ?

- Tiny signal as compared with the noise of our instruments and other signals from Earth, Solar System, Galaxy, which need to be accounted for

- Tiny signal as compared with the noise of our instruments and other signals from Earth, Solar System, Galaxy, which need to be accounted for

Resulting Strategy :

- Tiny signal as compared with the noise of our instruments and other signals from Earth, Solar System, Galaxy, which need to be accounted for

Resulting Strategy :

1. scan the sky with many detectors over the period of many years $\rightarrow$ lot of measurement

- Tiny signal as compared with the noise of our instruments and other signals from Earth, Solar System, Galaxy, which need to be accounted for

Resulting Strategy :

1. scan the sky with many detectors over the period of many years $\rightarrow$ lot of measurement
2. account for the noise $\rightarrow$ with statistical tools

- Tiny signal as compared with the noise of our instruments and other signals from Earth, Solar System, Galaxy, which need to be accounted for

Resulting Strategy :

1. scan the sky with many detectors over the period of many years $\rightarrow$ lot of measurement
2. account for the noise $\rightarrow$ with statistical tools
3. compress the data onto the map

- Tiny signal as compared with the noise of our instruments and other signals from Earth, Solar System, Galaxy, which need to be accounted for

Resulting Strategy :

1. scan the sky with many detectors over the period of many years → lot of measurement
2. account for the noise → with statistical tools
3. compress the data onto the map

Domain of time measurement : $n_t \approx O(10^{12-15})$ (following Moore's law)
Domain of the pixels : $n_p \approx O(10^{6+})$

What is map-making ?

Writting the vector of measurement $d \in \mathbb{R}^{n_t}$ as :

$$d = Ps + n \tag{1}$$

Where :

Writting the vector of measurement $d \in \mathbb{R}^{n_t}$ as :

$$d = Ps + n \tag{1}$$

Where :

- $s \in \mathbb{R}^{n_p}$ is the signal

Writting the vector of measurement $d \in \mathbb{R}^{n_t}$ as :

$$d = Ps + n \tag{1}$$

Where :

- $s \in \mathbb{R}^{n_p}$ is the signal
- $P \in \mathbb{R}^{n_t \times n_p}$ is the pointing matrix, tall and skinny and very sparse

Writting the vector of measurement $d \in \mathbb{R}^{n_t}$ as :

$$d = Ps + n \tag{1}$$

Where :

- $s \in \mathbb{R}^{n_p}$ is the signal
- $P \in \mathbb{R}^{n_t \times n_p}$ is the pointing matrix, tall and skinny and very sparse
- $n \in \mathbb{R}^{n_t}$ is the noise, modelled as a Gaussian stochastic process

Writting the vector of measurement $d \in \mathbb{R}^{n_t}$ as :

$$d = Ps + n \tag{1}$$

Where :

- $s \in \mathbb{R}^{n_p}$ is the signal
- $P \in \mathbb{R}^{n_t \times n_p}$ is the pointing matrix, tall and skinny and very sparse
- $n \in \mathbb{R}^{n_t}$ is the noise, modelled as a Gaussian stochastic process

Build the maximum likelihood estimate, $\hat{s}$, of the signal s given by :

Writting the vector of measurement $d \in \mathbb{R}^{n_t}$ as :

$$d = Ps + n \tag{1}$$

Where :

- $s \in \mathbb{R}^{n_p}$ is the signal
- $P \in \mathbb{R}^{n_t \times n_p}$ is the pointing matrix, tall and skinny and very sparse
- $n \in \mathbb{R}^{n_t}$ is the noise, modelled as a Gaussian stochastic process

Build the maximum likelihood estimate, $\hat{s}$, of the signal s given by :

$$\hat{s} = \left(P^t N^{-1} P\right)^{-1} P^t N^{-1} d \tag{2}$$

Where $N \in \mathbb{R}^{n_t \times n_t}$ is the covariance matrix of the noise.

Map-making : scanning strategy & maximum likelihood

What's a scanning strategy ?

Map-making : scanning strategy & maximum likelihood

What's a scanning strategy ?

The way we observe the sky, encoded by the pointing matrix P as such :
A line numbered $1 \leq i \leq n_t$ of P, $P_{i,.} \in \mathbb{R}^{n_p}$, says what pixels we look at time i

$$P_{i,.} = (0, ..., 0, t_i, 0, ..., 0) \tag{3}$$

Map-making : scanning strategy & maximum likelihood

What's a scanning strategy ?

The way we observe the sky, encoded by the pointing matrix P as such :
A line numbered $1 \leq i \leq n_t$ of P, $P_{i,.} \in \mathbb{R}^{n_p}$, says what pixels we look at time i

$$P_{i,.} = (0, ..., 0, t_i, 0, ..., 0) \tag{3}$$

Particular case : when polarization added, pixel domain*3 and lines of $P \in \mathbb{R}^{n_t \times 3n_p}$ became :

$$P_{i,.} = (0, ..., 0, t_i, q_i, u_i, 0, ..., 0) \tag{4}$$

What does N^{-1} looks like ?

What does N^{-1} looks like ?

N^{-1} is a block diagonal matrix

What does N^{-1} looks like ?

N^{-1} is a block diagonal matrix , many blocks per detector (i.e. many blocks)

What does N^{-1} looks like ?

N^{-1} is a block diagonal matrix , many blocks per detector (i.e. many blocks) , and each block is band-diagonal, toeplitz and symetric.

What does N^{-1} looks like ?

N^{-1} is a block diagonal matrix , many blocks per detector (i.e. many blocks) , and each block is band-diagonal, toeplitz and symetric.
Let's call :

- n_{det} : number of detector
- N_l^{-1} for $1 \leq l \leq n_{det}$ the blocks of N^{-1}
- d_l the diagonal coefficient of block l, and e_k^l for $2 \leq k \leq \lambda_l$ the off-diagonal coef. of block l, λ_l being the band width

What does N^{-1} looks like ?

N^{-1} is a block diagonal matrix , many blocks per detector (i.e. many blocks) , and each block is band-diagonal, toeplitz and symetric.
Let's call :

- n_{det} : number of detector
- N_l^{-1} for $1 \leq l \leq n_{det}$ the blocks of N^{-1}
- d_l the diagonal coefficient of block l, and e_k^l for $2 \leq k \leq \lambda_l$ the off-diagonal coef. of block l, λ_l being the band width

N^{-1} looks like :

$$N^{-1} = \begin{bmatrix} N_1^{-1} & 0 & \cdots & 0 \\ 0 & N_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & N_{n_{det}}^{-1} \end{bmatrix} \tag{5}$$

with block like this :

$$N_l^{-1} = \begin{bmatrix} d_l & e_2^l & \cdots & e_{\lambda_l}^l & 0 & \cdots & 0 \\ & \ddots & \ddots & & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & & \ddots & 0 \\ & & & d_l & e_2^l & & e_{\lambda_l}^l \\ & & * & & \ddots & \ddots & \vdots \\ & & & & & \ddots & e_2^l \\ & & & & & & d_l \end{bmatrix} \tag{6}$$

A few reminders on CG

Solve $A\underline{x} = b$, $A \in \mathbb{R}^{n \times n}$ SPD, $b, \underline{x} \in \mathbb{R}^n$.

Solve $A\underline{x} = b$, $A \in \mathbb{R}^{n \times n}$ SPD, $b, \underline{x} \in \mathbb{R}^n$.
For $x_0 \in \mathbb{R}^n$, build the sequence $(x_k)_{k \geq 0}$ s.t. :

$$\begin{cases} x_{k+1} \in x_0 + K_k \\ r_{k+1} := (b - Ax_{k+1}) \perp K_k \end{cases} \tag{7}$$

$K_k = \text{Span}(r_0, Ar_0, \ldots, A^{k-1}r_0)$, $r_0 = b - Ax_0$

Solve $A\underline{x} = b$, $A \in \mathbb{R}^{n\times n}$ SPD, $b, \underline{x} \in \mathbb{R}^n$.
For $x_0 \in \mathbb{R}^n$, build the sequence $(x_k)_{k\geq 0}$ s.t. :

$$\begin{cases} x_{k+1} \in x_0 + K_k \\ r_{k+1} := (b - Ax_{k+1}) \perp K_k \end{cases} \tag{7}$$

$K_k = \text{Span}(r_0, Ar_0, ..., A^{k-1}r_0)$, $r_0 = b - Ax_0$

Lemma

For x_k the k-th approximation build with (7), x_{k+1} satisfies :

$$||x_{k+1} - \underline{x}||_A = \min_{x\in x_0+K_k} ||x - \underline{x}||_A$$

Building $(x_k)_{k\geq 0}$ as :

$$x_{k+1} = x_k + \alpha_k p_k \tag{8}$$

Building $(x_k)_{k \geq 0}$ as :

$$x_{k+1} = x_k + \alpha_k p_k \tag{8}$$

Several choices of p_k $\rightarrow$ several CG-algorithm

Building $(x_k)_{k\geq 0}$ as :

$$x_{k+1} = x_k + \alpha_k p_k \tag{8}$$

Several choices of p_k $\rightarrow$ several CG-algorithm

ORTHOMIN	ORTHODIR
$p_{k+1} = r_{k+1} + \alpha_k p_k$ $\alpha_k = \frac{\lVert r_{k+1}\rVert_2}{\lVert r_k\rVert_2}$	$p_{k+1} = Ap_k - \gamma_k p_k - \sigma_k p_{k-1}$ $\gamma_k = (Ap_k)^t(Ap_k)$, $\sigma_k = (Ap_{k-1})^t(Ap_k)$

Algorithm 1 Conjugate Gradient

Require: $A, b, x_0 \in \mathbb{R}^n, \varepsilon > 0$

Ensure: $||Ax_{k+1} - b||_2 \leq \varepsilon ||b||_2$

$k = 0$

$p_0 = r_0 = Ax_0 - b$ or $p_0 = 0,\ p_1 = r_0 = Ax_0 - b$

while $||r_{k+1}|| > \varepsilon ||b||$ **do**

$\alpha_k = ||r_k||_2^2 / \left(p_k^t A p_k\right)$

$x_{k+1} = x_k - \alpha_k p_k$

$r_{k+1} = r_k - \alpha_k A p_k$

Build p_{k+1} with OMIN or ODIR

$k = k + 1$

end while

Return x_{k+1}

Enlarged CG

Define a splitting operator T_t, for $t \in \mathbb{N}$ the splitting parameter :

$$T_t : \begin{array}{c} \mathbb{R}^n \to \mathbb{R}^{n \times t} \\ x \mapsto T_t(x) \end{array} \tag{9}$$

with $T_t(x)$ s.t. $T_t(x) * \mathbf{1}_t = x$, $\mathbf{1}_t = (1)_{1 \leq j \leq t} \in \mathbb{R}^t$

Define a splitting operator T_t, for $t \in \mathbb{N}$ the splitting parameter :

$$T_t : \begin{array}{c} \mathbb{R}^n \to \mathbb{R}^{n \times t} \\ x \mapsto T_t(x) \end{array} \tag{9}$$

with $T_t(x)$ s.t. $T_t(x) * \mathbf{1}_t = x$, $\mathbf{1}_t = (1)_{1 \le j \le t} \in \mathbb{R}^t$

$$x = \begin{pmatrix} * \\ * \\ * \\ \vdots \\ * \\ * \\ * \\ * \end{pmatrix} \mapsto T_t(x) = \begin{pmatrix} * & 0 & \\ \vdots & \vdots & \\ * & 0 & \\ 0 & * & \\ \vdots & \vdots & \dots \\ \vdots & * & \\ \vdots & \vdots & \\ 0 & 0 & \end{pmatrix}$$

Enlarged the Krylov subspace with T_t :

$$K_{k,t} = \text{Span}_{\square}\left(T_t(r_0), AT_t(r_0), ..., A^{k-1}T_t(r_0)\right) \tag{10}$$

Enlarged the Krylov subspace with T_t :

$$K_{k,t} = \text{Span}_{\square}\left(T_t(r_0), AT_t(r_0), \ldots, A^{k-1}T_t(r_0)\right) \qquad (10)$$

For $x_0 \in \mathbb{R}^n$, build the sequence $(x_k)_{k\geq 0}$ s.t. :

$$\begin{cases} x_{k+1} \in x_0 + K_{k,t} \\ r_{k+1} := (b - Ax_{k+1}) \perp K_{k,t} \end{cases} \qquad (11)$$

Enlarged the Krylov subspace with T_t :

$$K_{k,t} = \text{Span}_{\square}\left(T_t(r_0), AT_t(r_0), ..., A^{k-1}T_t(r_0)\right) \qquad (10)$$

For $x_0 \in \mathbb{R}^n$, build the sequence $(x_k)_{k \geq 0}$ s.t. :

$$\begin{cases} x_{k+1} \in x_0 + K_{k,t} \\ r_{k+1} := (b - Ax_{k+1}) \perp K_{k,t} \end{cases} \qquad (11)$$

Lemma

For x_k the k-th approximation build with (11), x_{k+1} satisfies :

$$K_k \subset K_{k,t}$$

$$||x_{k+1} - \underline{x}||_A = \min_{x \in x_0 + K_{k,t}} ||x - \underline{x}||_A$$

Algorithm 2 Enlarged CG

Require: $A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n, x_0 \in \mathbb{R}^n, k_{max} \in \mathbb{N}, \varepsilon > 0$

Ensure: $||b - Ax_k||_2 < \varepsilon ||b||_2$ or $k = k_{max}$

$k = 0,\ p_0 = r_0 = b - Ax_0$

$X_0 = T_t(x_0),\ P_0 = T_t(p_0),\ R_0 = T_t(r_0)$

while $||r_{k+1}|| > \varepsilon ||b||$ ou $k < k_{max}$ **do**

Change P_{k+1} s.t. $P_{k+1}^t A P_{k+1} = \mathrm{Id}$

$\alpha_k = P_{k+1}^t R_k$

$X_{k+1} = X_k + P_{k+1}\alpha_k$

$R_{k+1} = R_k - AP_{k+1}\alpha_k$

$r_{k+1} = R_{k+1}\mathbf{1}_t$

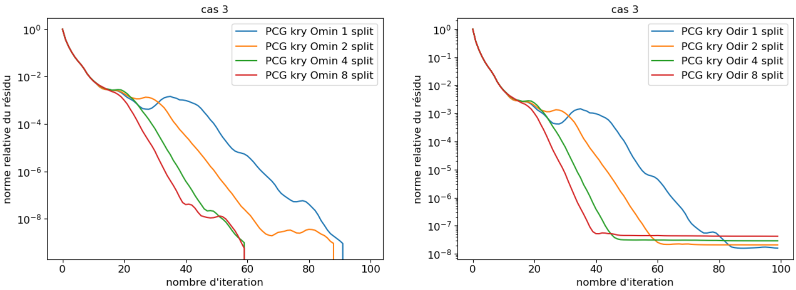
Build P_{k+1} with OMIN or ODIR

$k = k + 1$

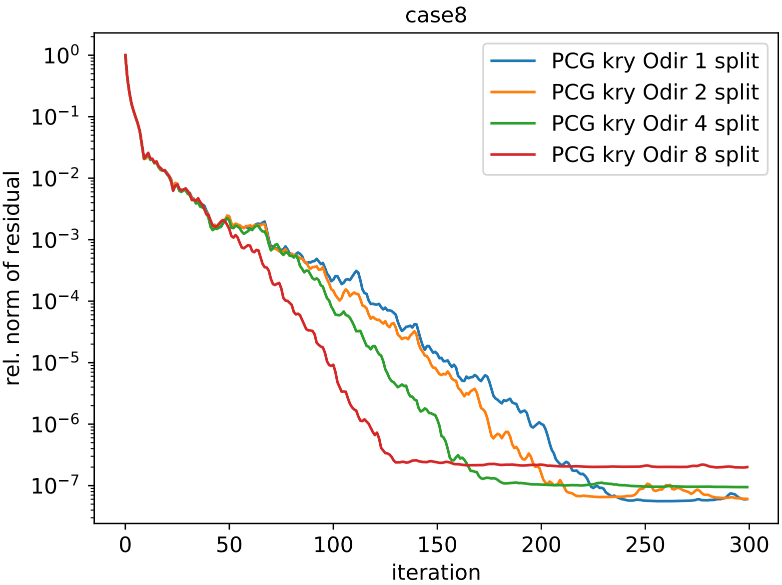
end while

Return $x_{k+1} = X_{k+1} * \mathbf{1}_t$

Numerical results


cas 3
norme relative du résidu
nombre d'iteration
PCG kry Omin 1 split
PCG kry Omin 2 split
PCG kry Omin 4 split
PCG kry Omin 8 split
cas 3
norme relative du résidu
nombre d'iteration
PCG kry Odir 1 split
PCG kry Odir 2 split
PCG kry Odir 4 split
PCG kry Odir 8 split

Case 3						
meth	Omin			Odir		
split	Iter	time/iter	total time	iter	time/iter	total time
1	65	28.512s	0.51h	65	26.156s	0.45h
2	50	63.419s	0.88h	50	49.207s	0.68h
4	39	124.642s	1.35h	39	100.946s	1.09h
8	34	226.933s	2.14h	34	201.005s	1.89h
32	25	874.021s	6.06h	25	779.239s	5.41h


case8
PCG kry Odir 1 split
PCG kry Odir 2 split
PCG kry Odir 4 split
PCG kry Odir 8 split
rel. norm of residual
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
0
50
100
150
200
250
300
iteration

What are the next steps ?

- Build effective preconditioner

- Build effective preconditioner
  - $M^{-1} = P^t diag(N^{-1})^{-1} P$ works as 1st lvl preconditioner, i.e. largest eigenvalue bounded by 2.

- Build effective preconditioner
  - $M^{-1} = P^t diag(N^{-1})^{-1} P$ works as 1st lvl preconditioner, i.e. largest eigenvalue bounded by 2.
  - More complex 2-lvl preconditioner from Domain Decomposition Methods (DMM) : GenEO

- Build effective preconditioner
  - $M^{-1} = P^t diag(N^{-1})^{-1} P$ works as 1st lvl preconditioner, i.e. largest eigenvalue bounded by 2.
  - More complex 2-lvl preconditioner from Domain Decomposition Methods (DMM) : GenEO
- Explore a lead to do the theory for block methods, especially Enlarged Krylov methods