Enlarged Krylov methods and 2-level preconditioner for the map-making problem in CMB data analysis

Modelling 2019 - Olomouc

Laura Grigori, Thibault Cimic

December 4, 2019

Laboratoire INRIA - Equipe ALPINES

The Cosmic Microwave Background and MM problem

The Cosmic Microwave Background

CMB : Cosmic Microwave Background

The Cosmic Microwave Background

CMB : Cosmic Microwave Background

Other names ?

The Cosmic Microwave Background

CMB : Cosmic Microwave Background

Other names ?

Relic radiation

The Cosmic Microwave Background

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)

The Cosmic Microwave Background

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)

The Cosmic Microwave Background

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)

The Cosmic Microwave Background

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 Cosmic Microwave Background

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 map of temperature and polarisation of these early photons

The Cosmic Microwave Background

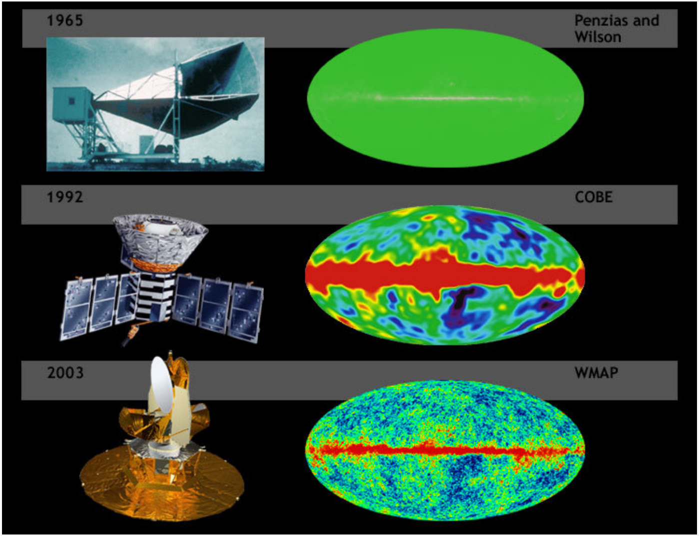

Figure 1: Evolution of the CMB map of temperature

The Cosmic Microwave Background

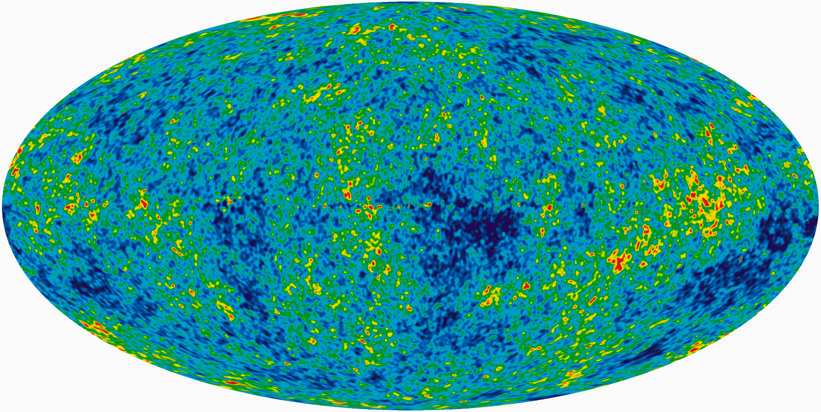

Figure 2: Map of temperature reconstructed from nine years of WMAP data satellite (2003-2012)

The Cosmic Microwave Background

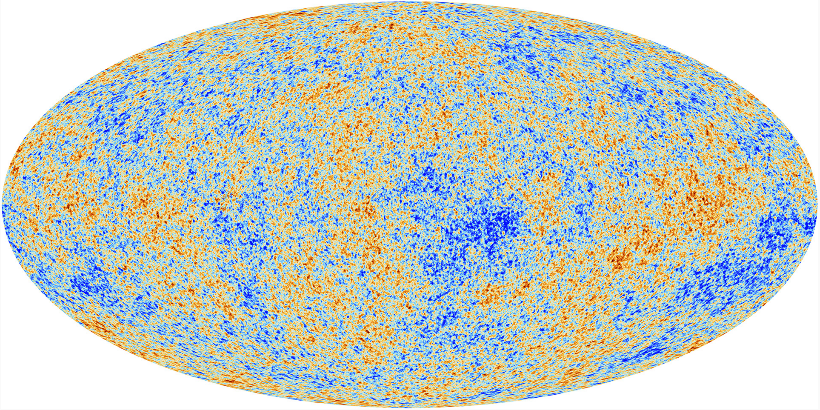

Figure 3: Map of temperature with Planck satellite

The map-making problem

The Map-making problem

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

$$d = Ps + n$$

The Map-making problem

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

$$d = Ps + n$$

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

The Map-making problem

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

$$d = Ps + n$$

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

The Map-making problem

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

$$d = Ps + n$$

- $s \in \mathbb{R}^{n_p}$ 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

The Map-making problem

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

$$d = Ps + n$$

- $s \in \mathbb{R}^{n_p}$ 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

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

$$\underbrace{\left(P^t N^{-1} P\right)}_{A} \hat{s} = P^t N^{-1} d$$

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

The Map-making problem

Map-making : scanning strategy & maximum likelihood

What's a scanning strategy ?

The Map-making problem

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

$$P_{i,.} = (0, ..., 0, t_i, 0, ..., 0) \tag{1}$$

The Map-making problem

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

$$P_{i,.} = (0, ..., 0, t_i, 0, ..., 0) \tag{1}$$

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

$$P_{i,.} = (0, ..., 0, t_i, q_i, u_i, 0, ..., 0) \tag{2}$$

The Map-making problem

What does N^{-1} looks like ?

The Map-making problem

What does N^{-1} looks like ?

N^{-1} is a block diagonal matrix, many blocks, and each block is band-diagonal, toeplitz and symetric.

The Map-making problem

What does N^{-1} looks like ?

N^{-1} is a block diagonal matrix, many blocks, and each block is band-diagonal, toeplitz and symetric.

Let's call :

- n_{blc} : number of blocks
- N_l^{-1} for $1 \leq l \leq n_{blc}$ 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

The Map-making problem

What does N^{-1} looks like ?

N^{-1} is a block diagonal matrix, many blocks, and each block is band-diagonal, toeplitz and symetric.

Let's call :

- n_{blc} : number of blocks
- N_l^{-1} for $1 \leq l \leq n_{blc}$ 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 :

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

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

with block like this :

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

From CG to Enlarged-CG

The Enlarged-CG

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

$$T_t : \begin{array}{c} \mathbb{R}^n \to \mathbb{R}^{n \times t} \\ x \mapsto T_t(x) \end{array} \tag{3}$$

with $T_t(x)$ s.t. $T_t(x) * \mathbf{1}_t = x$ and $T_t(x)$ has t linearly independent columns.

The Enlarged-CG

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

$$T_t : \begin{array}{c} \mathbb{R}^n \rightarrow \mathbb{R}^{n\times t} \\ x \mapsto T_t(x) \end{array} \tag{3}$$

with $T_t(x)$ s.t. $T_t(x) * \mathbf{1}_t = x$ and $T_t(x)$ has t linearly independent columns.

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

The Enlarged-CG

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

The Enlarged-CG

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

Enlarged the Krylov space with T_t :

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

The Enlarged-CG

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

Enlarged the Krylov space with T_t :

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

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

$$\begin{cases} x_{k+1} \in x_0 + K_{k,t} \\ r_{k+1} = b - Ax_k \perp K_{k,t} \end{cases} \tag{4}$$

Lemma

For x_k the k-th approximation build from (4), x_{k+1} satisfies :

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

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

Theorem (O. Tissot, L. Grigori)

Let x_k be the k-th iterate build with (10), then we have :

$$||x_k - \underline{x}||_A \leq C \left(\frac{\sqrt{\kappa_t} - 1}{\sqrt{\kappa_t} + 1} \right)^k \quad (5)$$

with $\kappa_t = \lambda_n / \lambda_t$ where λ_t is the t^{th} smallest eigenvalue of A and C is a constant independent of k.

Algorithm 1 Enlarged CG

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

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**

A-orthonormalize P_k

$\alpha_k = P_k^t R_k$

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

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

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

$P_{k+1} = R_{k+1} - P_k P_k^\top AR_{k+1}$

$k = k + 1$

end while

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

A few numerical results

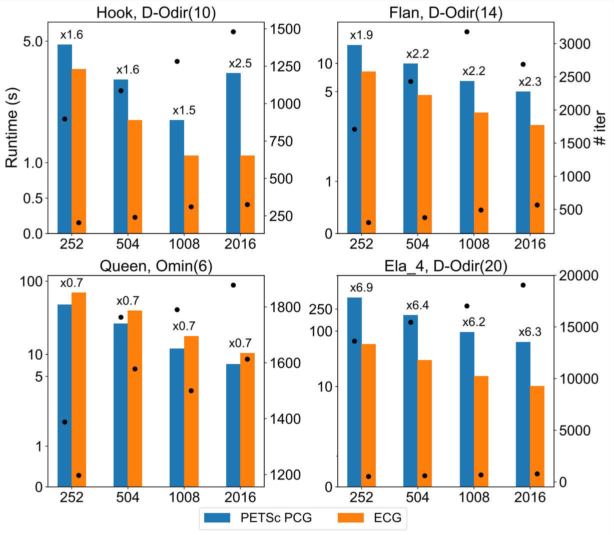

Figure 4: Tim Davis' collection + 3D elasticity problem, O. Tissot

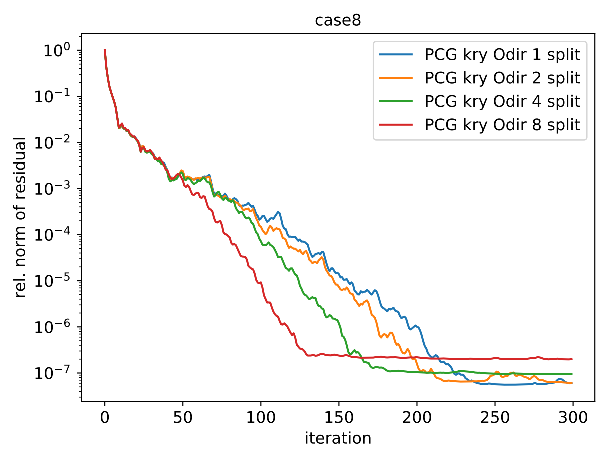

Figure 5: Map-Making ORTHODIR case8

2-level preconditioner from fictitious space lemma

The fictitious space lemma

Lemma

$(H, (.,.))$, $(H_D, (.,.)_D)$ two Hilbert spaces, two symmetric positive bilinear forms $a : H \times H \longrightarrow \mathbb{R}$, $b : H_D \times H_D \longrightarrow \mathbb{R}$, generated by the SPD operators $\mathcal{A} : H \longrightarrow H$ and $B : H_D \longrightarrow H_D$, respectively. Suppose that there exists a linear operator $\mathcal{R} : H_D \longrightarrow H$ such that the following holds :

- *$\mathcal{R}$ is surjective.*
- $\exists c_U$ *s.t.* $\forall u_D \in H_D, a(\mathcal{R}u_D, \mathcal{R}u_D) \leq c_U b(u_D, u_D)$
- $\exists c_L$ *s.t.* $\forall u \in H, \exists u_D \in H_D$ *s.t.* $\mathcal{R}u_D = u$, $c_L b(u_D, u_D) \leq a(\mathcal{R}u_D, \mathcal{R}u_D) = a(u, u)$

$\mathcal{R}^* : H \longrightarrow H_D$ the adjoint operator of $\mathcal{R}$, then :

$$\Lambda\left(\mathcal{R}B^{-1}\mathcal{R}^*\mathcal{A}\right) \subset [c_L, c_U]$$

The fictitious space lemma

- n_{blc} blocks of N^{-1} splits $\{1, ..., n_t\}$ in n_{blc} domains :

$$\mathbb{R}^{n_t} \cong \bigcup_{l=0}^{n_{blc}} \mathbb{R}^{n_l} =: H_D$$

The fictitious space lemma

- n_{blc} blocks of N^{-1} splits $\{1, ..., n_t\}$ in n_{blc} domains :

$$\mathbb{R}^{n_t} \cong \bigcup_{l=0}^{n_{blc}} \mathbb{R}^{n_l} =: H_D$$

$$P^\top = (P_1^\top | ... | P_{n_{blc}}^\top) \text{ and } P^\top N^{-1} P = \sum_{l=1}^{n_{blc}} P_l^\top N_l^{-1} P_l$$

The fictitious space lemma

- n_{blc} blocks of N^{-1} splits $\{1, ..., n_t\}$ in n_{blc} domains :

$$\mathbb{R}^{n_t} \cong \bigcup_{l=0}^{n_{blc}} \mathbb{R}^{n_l} =: H_D$$

$$P^\top = (P_1^\top|...|P_{n_{blc}}^\top) \text{ and } P^\top N^{-1} P = \sum_{l=1}^{n_{blc}} P_l^\top N_l^{-1} P_l$$

- Define :

$$\mathcal{R} : \begin{array}{c} H_D \longrightarrow H \\ (u_l)_{l=0...N} \mapsto \sum_{l=0}^{N} P_l^\top u_l \end{array}$$

The fictitious space lemma

- n_{blc} blocks of N^{-1} splits $\{1, ..., n_t\}$ in n_{blc} domains :

$$\mathbb{R}^{n_t} \cong \bigcup_{l=0}^{n_{blc}} \mathbb{R}^{n_l} =: H_D$$

$P^\top = (P_1^\top|...|P_{n_{blc}}^\top)$ and $P^\top N^{-1} P = \sum_{l=1}^{n_{blc}} P_l^\top N_l^{-1} P_l$

- Define :

$$\mathcal{R} : \begin{array}{c} H_D \longrightarrow H \\ (u_l)_{l=0...N} \mapsto \sum_{l=0}^{N} P_l^\top u_l \end{array}$$

then :

$$\mathcal{R}^* : \begin{array}{c} H \longrightarrow H_D \\ u \mapsto (P_i u)_{i=0...N} \end{array}$$

The fictitious space lemma

- n_{blc} blocks of N^{-1} splits $\{1, ..., n_t\}$ in n_{blc} domains :

$$\mathbb{R}^{n_t} \cong \bigcup_{l=0}^{n_{blc}} \mathbb{R}^{n_l} =: H_D$$

$P^\top = (P_1^\top|...|P_{n_{blc}}^\top)$ and $P^\top N^{-1} P = \sum_{l=1}^{n_{blc}} P_l^\top N_l^{-1} P_l$

- Define :

$$\mathcal{R} : \begin{array}{c} H_D \longrightarrow H \\ (u_l)_{l=0...N} \mapsto \sum_{l=0}^{N} P_l^\top u_l \end{array}$$

then :

$$\mathcal{R}^* : \begin{array}{c} H \longrightarrow H_D \\ u \mapsto (P_i u)_{i=0...N} \end{array}$$

Lemma (Surjectivity of $\mathcal{R}$)

$\mathcal{R}$ define as such is surjective.

The fictitious space lemma

- Define operators a and b :

$$a : \begin{array}{c} H \times H \longrightarrow \mathbb{R} \\ (u, v) \mapsto a(u, v) := u^\top A v \end{array}$$

$$b : \begin{array}{c} H_D \times H_D \longrightarrow \mathbb{R} \\ (\mathcal{U}, \mathcal{V}) \mapsto b(\mathcal{U}, \mathcal{V}) := \mathcal{U}^\top B \mathcal{V} \end{array}$$

The fictitious space lemma

- Define operators a and b :

$$a: \begin{array}{c} H \times H \longrightarrow \mathbb{R} \\ (u,v) \mapsto a(u,v) := u^\top A v \end{array}$$

$$b: \begin{array}{c} H_D \times H_D \longrightarrow \mathbb{R} \\ (\mathcal{U},\mathcal{V}) \mapsto b(\mathcal{U},\mathcal{V}) := \mathcal{U}^\top B \mathcal{V} \end{array}$$

$$B(\mathcal{U}) := \left(P_0^\top A P_0 u_0, diag(\mathcal{N}_1^{-1}) u_1, ..., diag(\mathcal{N}_{n_{blc}}^{-1}) u_{n_{blc}}\right)$$

$$B^{-1}(\mathcal{U}) :=$$

$$\left((P_0^\top A P_0)^{-1} u_0, diag(\mathcal{N}_1^{-1})^{-1} u_1, ..., diag(\mathcal{N}_{n_{blc}}^{-1})^{-1} u_{n_{blc}}\right)$$

The fictitious space lemma

- Define operators a and b :

$$a : \begin{array}{c} H \times H \longrightarrow \mathbb{R} \\ (u, v) \mapsto a(u, v) := u^\top A v \end{array}$$

$$b : \begin{array}{c} H_D \times H_D \longrightarrow \mathbb{R} \\ (\mathcal{U}, \mathcal{V}) \mapsto b(\mathcal{U}, \mathcal{V}) := \mathcal{U}^\top B \mathcal{V} \end{array}$$

$$B(\mathcal{U}) := \left(P_0^\top A P_0 u_0, diag(\mathcal{N}_1^{-1}) u_1, ..., diag(\mathcal{N}_{n_{blc}}^{-1}) u_{n_{blc}}\right)$$

$$B^{-1}(\mathcal{U}) := \left((P_0^\top A P_0)^{-1} u_0, diag(\mathcal{N}_1^{-1})^{-1} u_1, ..., diag(\mathcal{N}_{n_{blc}}^{-1})^{-1} u_{n_{blc}}\right)$$

- Set the preconditioner M_2^{-1} :

$$M_2^{-1} = \mathcal{R} B^{-1} \mathcal{R}^* = P_0 (P_0^\top M P_0)^{-1} P_0^\top + P^\top diag(\mathcal{N}^{-1})^{-1} P$$

The fictitious space lemma

Lemma (Continuity of $\mathcal{R}$)

As such : $\exists c_U$ *s.t.* $\forall u_D \in H_D, a(\mathcal{R}u_D, \mathcal{R}u_D) \leq c_U b(u_D, u_D)$

The fictitious space lemma

Lemma (Continuity of $\mathcal{R}$)

As such : $\exists c_U$ *s.t.* $\forall u_D \in H_D, a(\mathcal{R}u_D, \mathcal{R}u_D) \leq c_U b(u_D, u_D)$

This result uses :

- for all $u_D \in H_D$:

$$u_D^\top N^{-1} u_D \leq K u_D^\top diag(N^{-1}) u_D$$

The fictitious space lemma

Coarse space correction :

- GenEO preconditioner : Neumann matrices $\tilde{A}^j$, uses PDE setting

$$\sum_{i=1}^{N} (R_j U)^\top \tilde{A}^j R_j U \leq k_1 U^\top A U$$

The fictitious space lemma

Coarse space correction :

- GenEO preconditioner : Neumann matrices $\tilde{A}^j$, uses PDE setting

$$\sum_{i=1}^{N} (R_j U)^\top \tilde{A}^j R_j U \leq k_1 U^\top A U$$

- ALS preconditioner : ALS of matrix A, $\tilde{A}^j$, uses operator R of DD

$$\sum_{i=1}^{N} U^\top \tilde{A}^j U \leq k_m U^\top A U$$

The fictitious space lemma

In map-making : $P^\top N^{-1} P = \sum_{i=1}^{n_{blc}} P_i^\top N_i^{-1} P_i$

$$u^\top P_i^\top N_i^{-1} P_i u \leq u^\top P^\top N^{-1} P u$$

The fictitious space lemma

In map-making : $P^\top N^{-1} P = \sum_{i=1}^{n_{blc}} P_i^\top N_i^{-1} P_i$

$$u^\top P_i^\top N_i^{-1} P_i u \leq u^\top P^\top N^{-1} P u$$

Ongoing/future work :

The fictitious space lemma

In map-making : $P^\top N^{-1} P = \sum_{i=1}^{n_{blc}} P_i^\top N_i^{-1} P_i$

$$u^\top P_i^\top N_i^{-1} P_i u \leq u^\top P^\top N^{-1} P u$$

Ongoing/future work :

- Describe the u_D s.t. $\mathcal{R} u_D = u$ for a given $u \in H$

The fictitious space lemma

In map-making : $P^\top N^{-1} P = \sum_{i=1}^{n_{blc}} P_i^\top N_i^{-1} P_i$

$$u^\top P_i^\top N_i^{-1} P_i u \leq u^\top P^\top N^{-1} P u$$

Ongoing/future work :

- Describe the u_D s.t. $\mathcal{R} u_D = u$ for a given $u \in H$
- Build a coarse space that will give stable decomposition property using generalized eigenvalue problem build in a similar way as DD or ALS.

The fictitious space lemma

In map-making : $P^\top N^{-1} P = \sum_{i=1}^{n_{blc}} P_i^\top N_i^{-1} P_i$

$$u^\top P_i^\top N_i^{-1} P_i u \leq u^\top P^\top N^{-1} P u$$

Ongoing/future work :

- Describe the u_D s.t. $\mathcal{R} u_D = u$ for a given $u \in H$
- Build a coarse space that will give stable decomposition property using generalized eigenvalue problem build in a similar way as DD or ALS.
- Implement those two methods on highly parallel architure.

References

(1) L. Grigori, S. Moufawad, F. Nataf : Enlarged Krylov subspace conjugate gradient methods for reducing communication, SIAM Journal on Matrix Analysis and Applications, 2016

(2) L. Grigori, O. Tissot : Scalable Linear Solvers based on Enlarged Krylov subspaces with Dynamic Reduction of Search Directions, Research Rapport RR-9190, 2018

(3) L. Grigori, H. Al Daas : A class of efficient locally constructed preconditioners based on coarse spaces, SIAM Journal on Matrix Analysis and Applications, 2019

(4) V. Dolean, P. Jolivet, F. Nataf : An introduction to domain decomposition methods, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2015

Thank you for you attention !