Trefftz variational iterative methods for solving linear hyperbolic systems

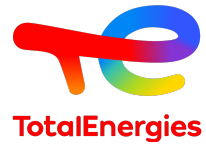
Sébastien Tordeux

Trefftz variational iterative method for solving linear hyperbolic systems

Sébastien Tordeux

EPI Makutu, INRIA, TotalEnergies, Université de Pau et des Pays de l'Adour, Bordeaux INP, CNRS

Workshop BIRS in Oaxaca, 2024 O5/17



Section 1

Introduction

Maxwell system

Find $\mathbf{E} \in L^2(\mathrm{rot}\,, \Omega)$ and $\mathbf{H} \in L^2(\mathrm{rot}\,, \Omega)$ such that

$$\left\{ \begin{array}{rcl} \nabla \times \mathbf{H} & = & ik\,\varepsilon_r \mathbf{E} \\ \nabla \times \mathbf{E} & = & -ik\,\mu_r \mathbf{H}, \text{ dans } \Omega, \end{array} \right.$$

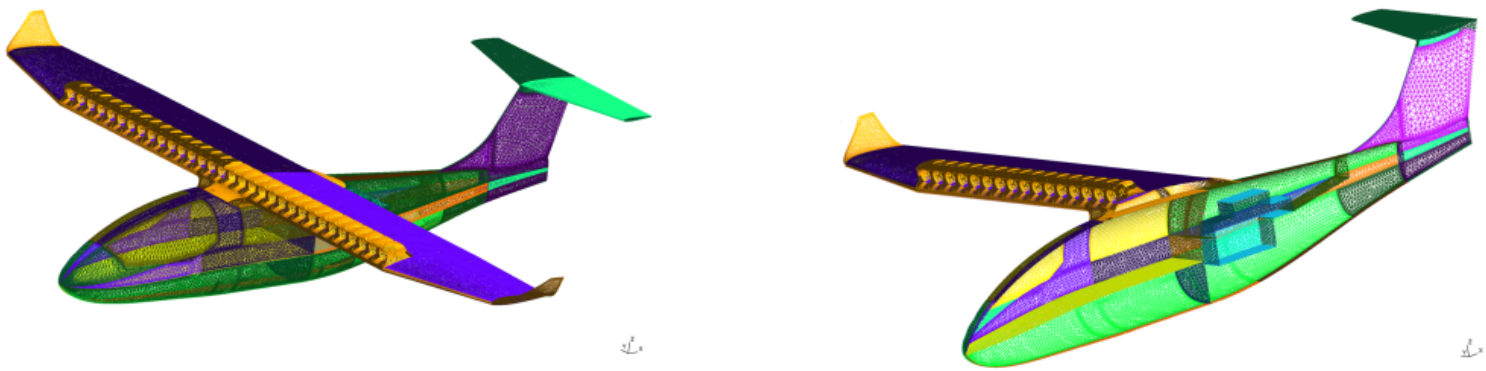
with boundary conditions

$$n \times \mathbf{H} - Z\, n \times (n \times \mathbf{E}) \;=\; \varphi \text{ sur } \partial\Omega.$$

where Ω is a 3D domain, ε_0 and μ_0 are piecewise constant.

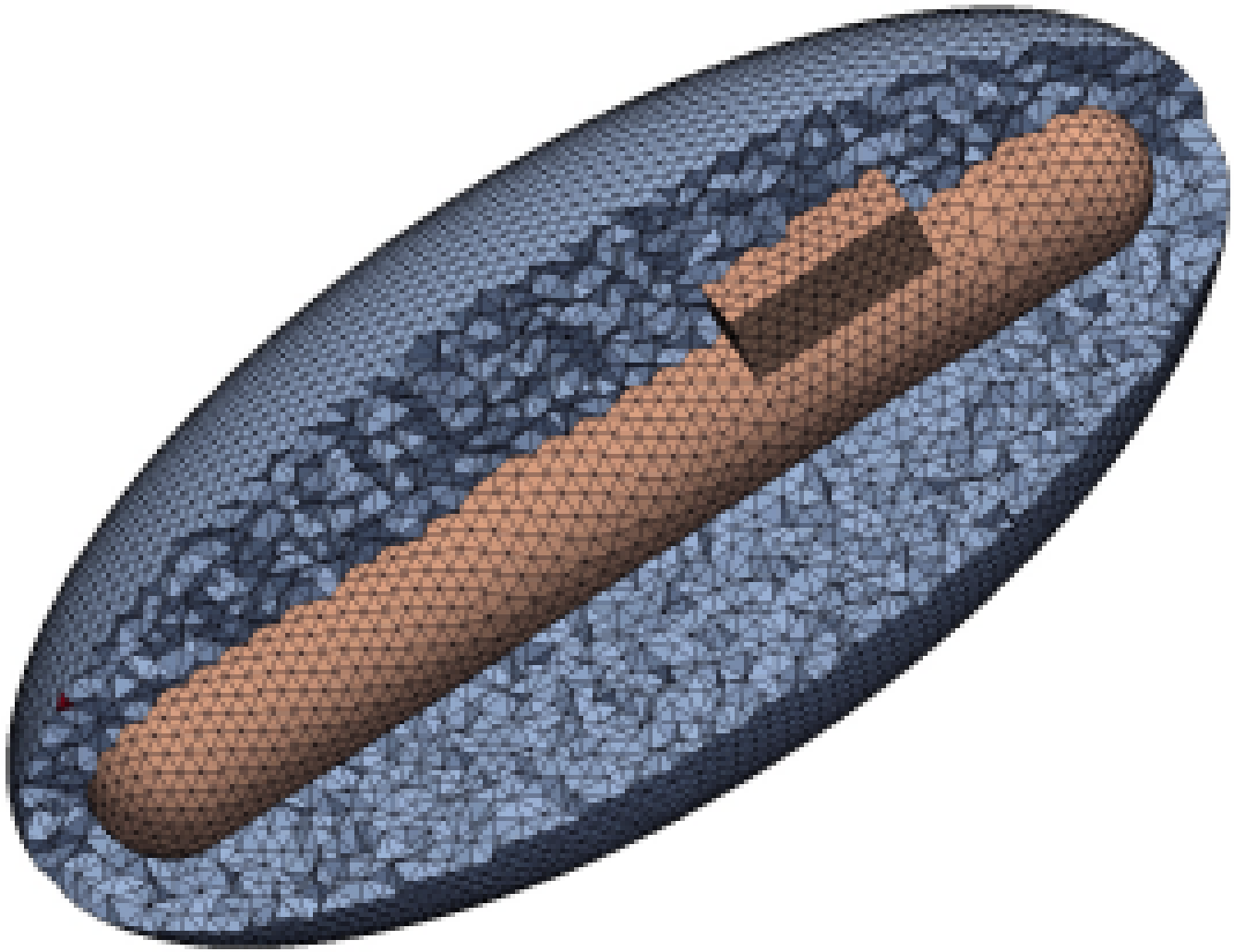
Objective:performing simulation in very large domains

$$\rightarrow \textbf{until } \mathbf{200} \text{ wavelengths} = \mathbf{200\lambda = 8 \times 10^6} \text{ elements.}$$



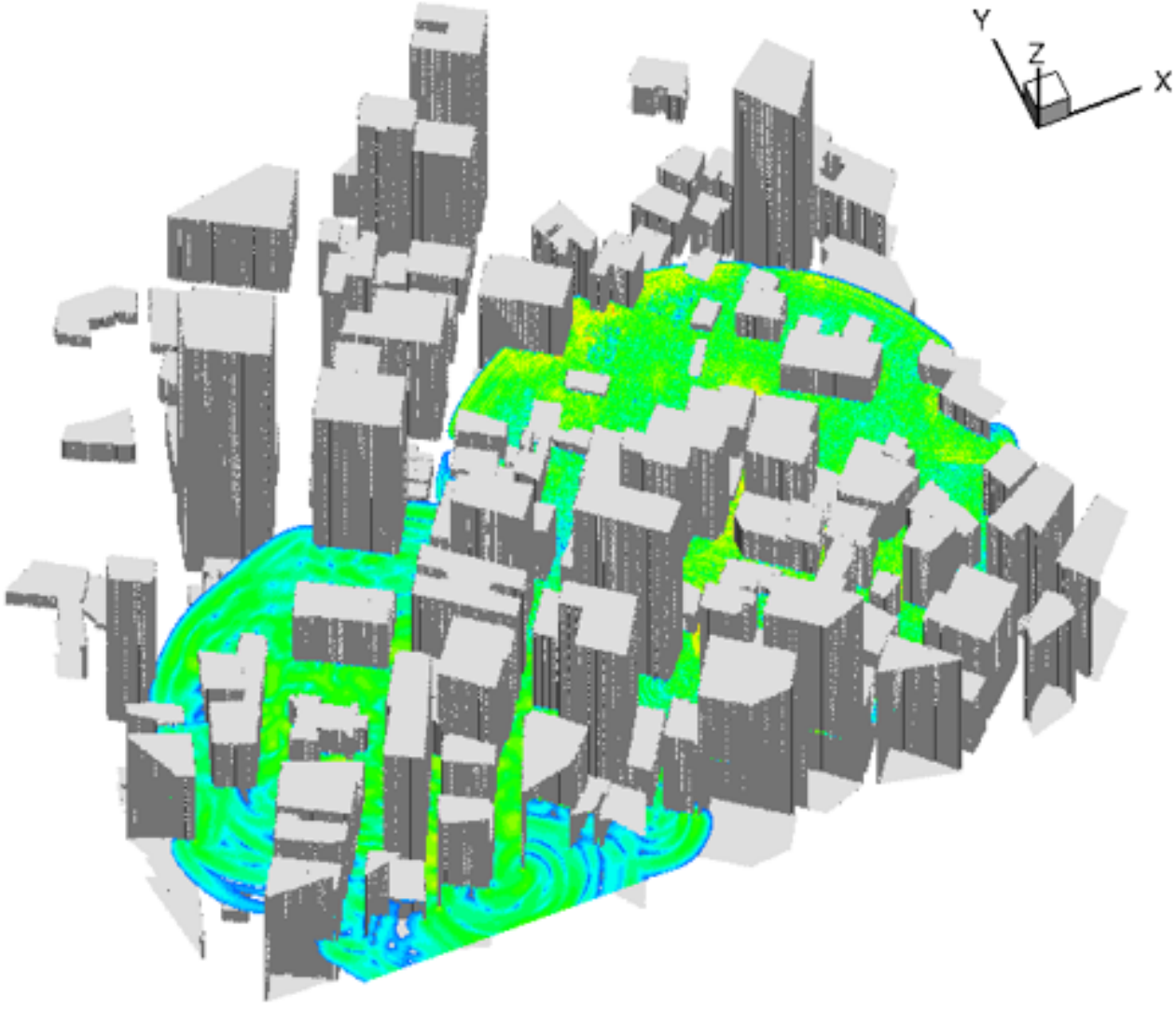
Geometry of Ampere prototype (DEMR ONERA).

Other examples of applications



Scattering by a submarine

Other examples of applications



Mesh of Manhattan by T. Volpert (ONERA)

Helmholtz equation

Helmholtz problem reads

$$\begin{cases} \Delta u + k^2 u = 0 & \text{dans } \Omega, \\ \frac{\partial u}{\partial n} - iku = \varphi & \text{sur } \partial\Omega. \end{cases}$$

Find $u \in H^1(\Omega)$ such that for any $v \in H^1(\Omega)$

$$\int_\Omega \nabla u \cdot \overline{\nabla v} - k^2 u\overline{v} - ik \int_{\partial\Omega} u\overline{v} = \int_{\partial\Omega} \varphi\overline{v}$$

From this weakly elliptic form, we easily get

- the well-posedness (Fredholm alternative, unique continuation)
- convergent finite element formulations

Is it the end of the story ?

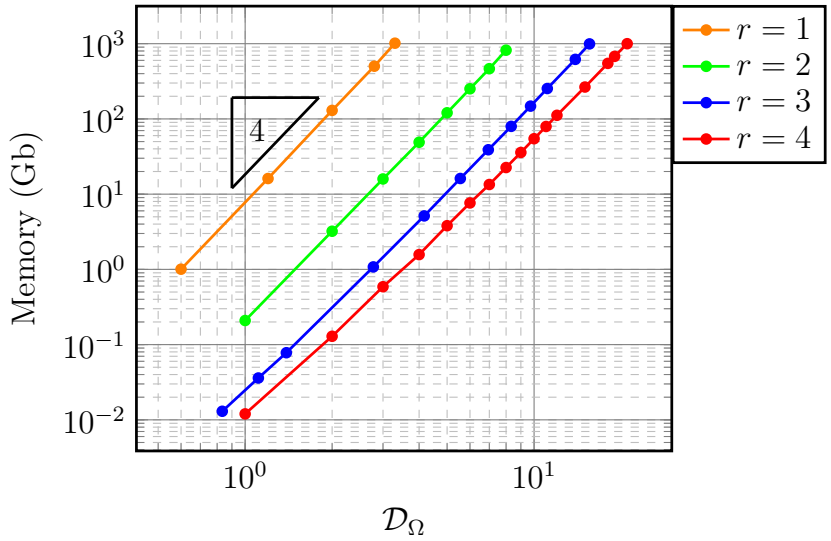
F. Ihlenburg and I. Babuska, *Finite element solution of the Helmholtz equation with high wave number part II: the hp version of the FEM, SIAM Journal on Numerical Analysis*, 34 (1), pp. 315–358 (1997).

Oliver G. Ernst, and Martin J. Gander. "Why it is difficult to solve Helmholtz problems with classical iterative methods." Numerical analysis of multiscale problems (2012): 325-363.

Memory limits of classical solvers

$$\mathbf{A}\,\mathbf{x} = \mathbf{F}$$

- A minimum number of unknowns:of second order
  - EDP
  - continuous finite elements
- Finite element methods handle geometry and evanescent modes very well.
- **Memory limitation**for LU methods.

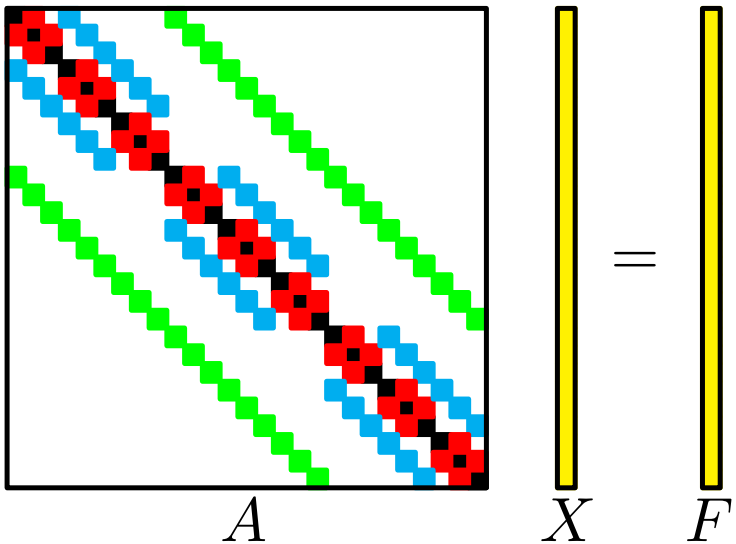


Memory cost (factorization LU of MUMPS©) for different sizes of domain $\mathcal{D}_\Omega$ and different approximation orders r.
Memory $= (\mathcal{D}_\Omega)^4$.

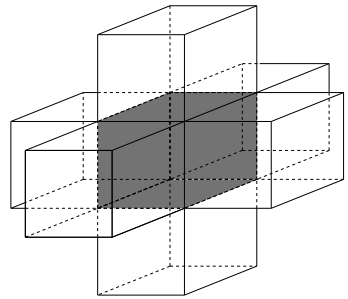
	Memory	Sockets	$\mathcal{D}_\Omega$	Power	Cost
PC	32 Gb	1	10λ	200 W	0.034 €/h
Parallel computing	1 Tb	16	24λ	3.2 KW	0.544 €/h
HPC	320 Tb	5000	100λ	1.0 MW	170 €/h
HPC+	5120 Tb	80000	200λ	16 MW	2720 €/h

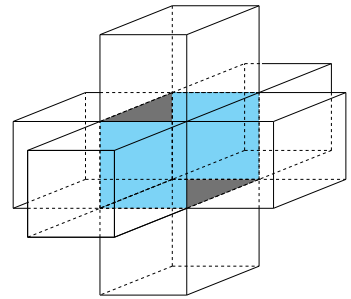
Matrix structure

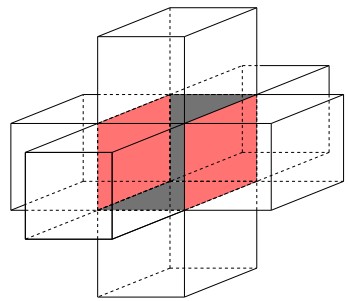
$$AX = F$$

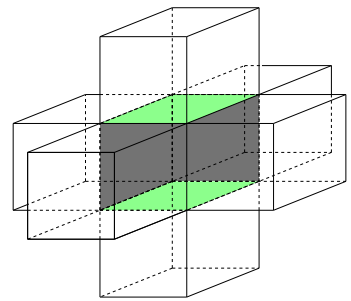


Matrix A for $27 = 3^3$ cubes.



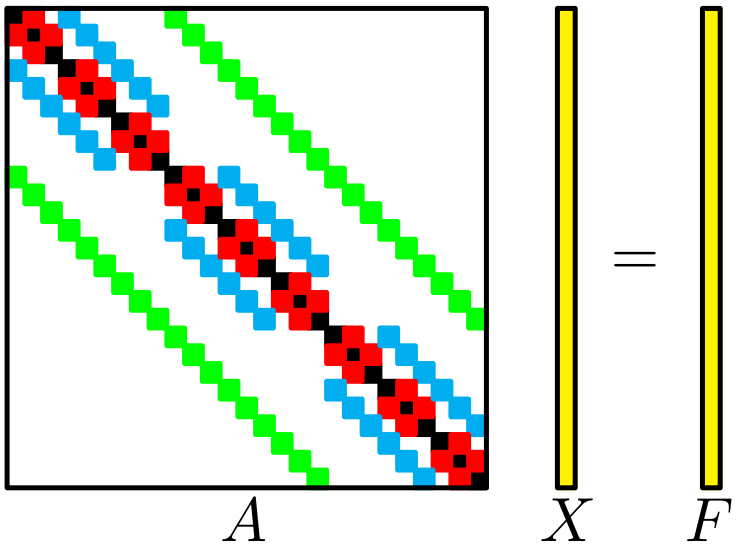






Matrix structure

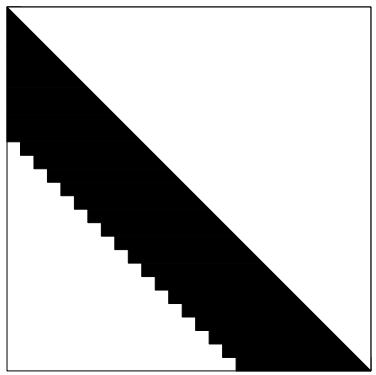
$AX = F$

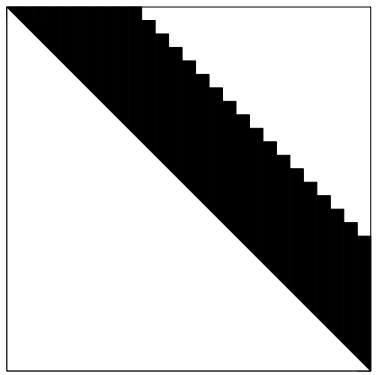


Matrix A for $27 = 3^3$ cubes.

$A = L\,U$

- L: lower triangular matrix
- U: upper triangular matrix





			L		U
Size of the domain	3λ	5λ	10λ	20λ	100λ
Number of elements	27	512	1000	8000	10^6
LU Memory	10 MB	135 MB	4.3 GB	138 GB	432 TB

Krylov methods

Cayleigh Hamilton theorem *If* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *is invertible, there exists a polynomial* p

$$\mathbf{A}^{-1} = p(\mathbf{A}) \text{ with } deg(p) < n.$$

Iterative Krylov method for solving $AX = F$

$$\mathbb{K}_k = \text{span}\Big(\Big\{\mathbf{F},\, \mathbf{AF},\, \mathbf{A}^2\mathbf{F},\, \ldots,\, \mathbf{A}^{k-1}\mathbf{F}\Big\}\Big).$$

Find $\mathbf{X}_k \in \mathbb{K}_{N_{\text{kry}}}$ that minimzes

$$J(\mathbf{X}_k) = \|\mathbf{AX}_k - \mathbf{F}\|_2^2$$

In theory, we have

- Convergence in n iterations $\mathbf{X}_n = \mathbf{X}$
- Error estimate for small k

$$\|\mathbf{F}\|_2 \geq \|\mathbf{AX}_1 - \mathbf{F}\|_2 \geq \cdots \geq \|\mathbf{AX}_k - \mathbf{F}\|_2 \geq \|\mathbf{AX}_{k+1} - \mathbf{F}\|_2$$

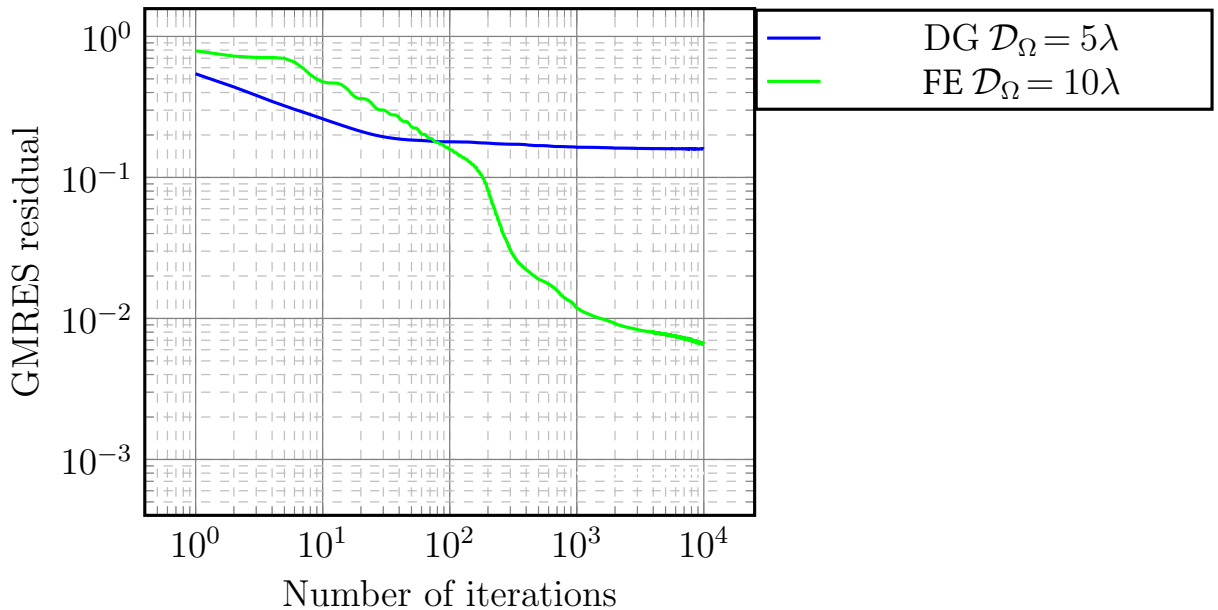
If the spectrum of A is included in $B(\lambda_0, r)$ with $r < |\lambda_0|$, then $\forall \varepsilon > 0$, $\exists C_\varepsilon > 0$

$$\|\mathbf{X}_k - \mathbf{X}\|_2 \;\leq\; C_\varepsilon \left(\frac{|\lambda|_{max} - |\lambda|_{min}}{|\lambda|_{max} + |\lambda|_{min}} + \varepsilon\right)^k \|\mathbf{F}\|_2$$

GMRES in use

Efficient implementation suggested by Saad and Schultz[1]

- GMRES methods are not exact !
- **GMRES methods are very slowly converging**



- We therefore prefer to use restarts to clear memory.

This is one of the curses of the large dimension.

[1] Saad, Youcef and Schultz, Martin H,GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems,SIAM,7,3,856–869,1986

There exist partial solutions

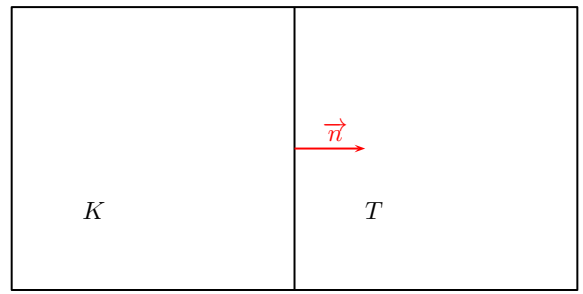
- Global Preconditionnings
  - → A. Vion and C. Geuzaine. Double sweep preconditioner for optimized Schwarz methods applied to the Helmholtz problem. J. Comput. Phys., 2014.
  - → M. Gander, L. Halpern, KS. L. Repiquet A new coarse grid correction for RAS/AS. In Domain Decomposition Methods in Science and Engineering XXI 2014.

- Domain decomposition methods

$$\begin{cases} Y\ p_T^{k+1} + i\kappa\nabla p_T^{k+1}\cdot\overrightarrow{n} = Y\ p_K^{k} + i\kappa\nabla p_K^{k}\cdot\overrightarrow{n}, \\ Y\ p_K^{k+1} - i\kappa\nabla p_K^{k+1}\cdot\overrightarrow{n} = Y\ p_T^{k} - i\kappa\nabla p_T^{k}\cdot\overrightarrow{n}, \end{cases}$$

→ A. Modave, X. Antoine et C. Geuzaine. An efficient domain decomposition method with cross-point treatment for Helmholtz problems. In CSMA 2019-14e Colloque National en Calcul des Structures. 2019.

→ F. Collino, P. Joly, and M. Lecouvez. "Exponentially convergent non overlapping domain decomposition methods for the Helmholtz equation." ESAIM: Math. Mod. and Num. An. 54.3. 775-810. 2020.

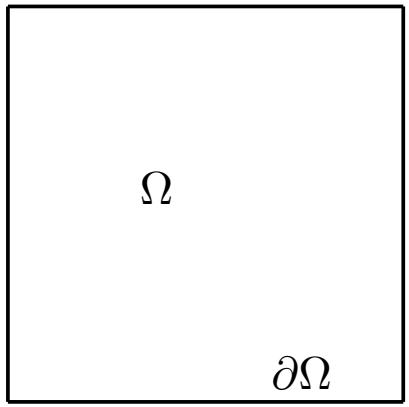


- Hyperbolic Trefftz methods
  - → Back to hyperboliticity
  - → Naturally adapted to GMRES methods

Section 2

Standard Trefftz methods

Classical variational Trefftz methods



The sample problem reads: find $p \in H^1(\Omega)$ such that

$$\begin{cases} \Delta p + k^2 p = 0 & \text{in } \Omega, \\ \frac{\partial p}{\partial n} - ikp = \varphi & \text{on } \partial\Omega. \end{cases}$$

with $\varphi \in L^2(\partial T)$.

Main features of the Trefftz space are

- it belongs to the discontinuous Galerkin spaces
- it is composed of local solutions to the PDE of interest

It is denoted by X and defined by

$$X = \left\{ p' : \Omega \longrightarrow \mathbb{C} \text{ such that } p'_K \in X_K \right\}$$

p'_K is the restriction of p' to element K and

$$X_K = \left\{ p'_K \in H^1(K) \text{ such that } \Delta p'_K + k^2 p'_K = 0, \quad \frac{\partial p'_K}{\partial n} \in L^2(\partial K) \right\}$$

Mesh of the domain Ω

- $K \in \mathcal{K}$ denotes one element of the mesh
- ∂K is the boundary of K
- $\mathcal{K}$ is the set of elements
- $\overline{\Omega} = \bigcup_{K \in \mathcal{K}} \overline{K}$.

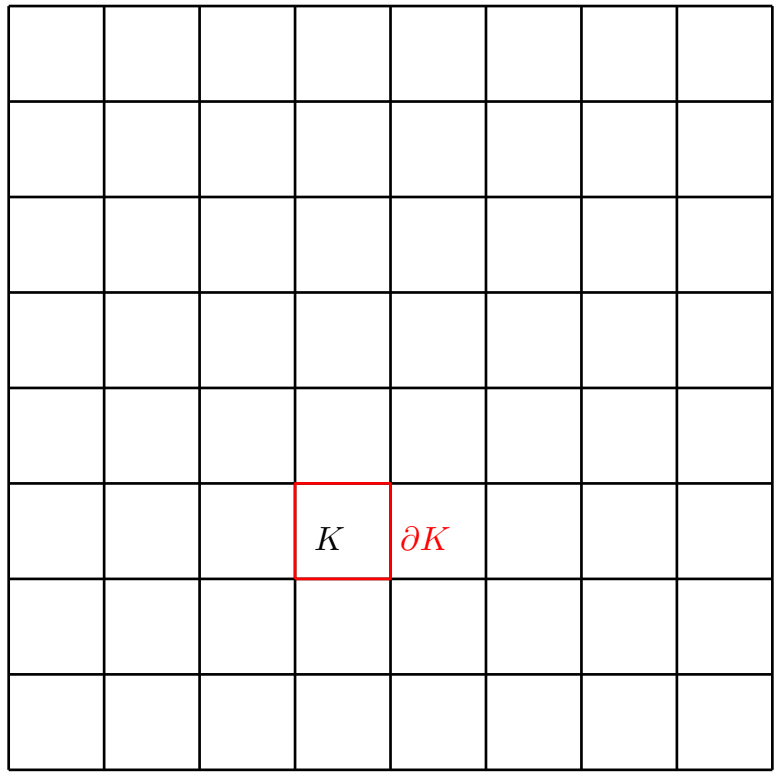
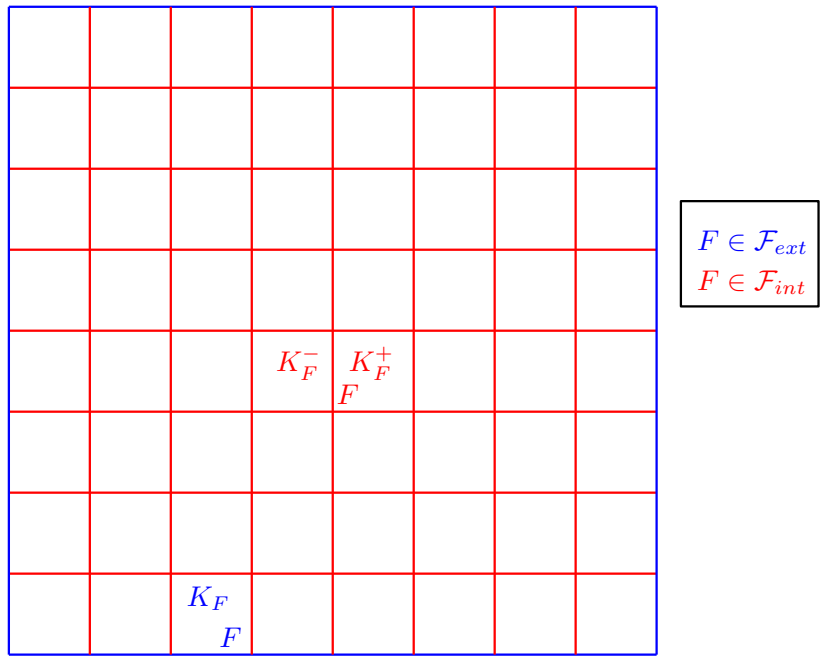


Figure: Domain discretization Ω.

Mesh of the domain Ω

$$\partial K = \bigcup_{F \in \mathcal{F}_K} F$$

- $\mathcal{F}_K$ denotes the set of the faces K
- F stands for a face
- $\mathcal{F}_{int}$ is the set of interior faces
- $\mathcal{F}_{ext}$ is the set of exterior faces



- If F is an interior face, then it hits two elements K_F^+ et K_F^-.
- If F is an exterior face, then it touches only one element K_F.

Reciprocity principle

Green formula reads:

$$\int_{\partial K} \frac{\partial p_K}{\partial n} \overline{p'_K} \;=\; \int_K \nabla p_K \overline{\nabla p'_K} - k^2\, p_K \overline{p'_K} \;=\; \int_{\partial K} p_K \overline{\frac{\partial p'_K}{\partial n}} \quad \text{for } p' \in X.$$

We sum up all the elements:

$$\sum_{K\in\mathcal{K}} \int_{\partial K} \nabla p_K \cdot \overrightarrow{n}_K \, \overline{p'_K} - p_K \, \overline{\nabla p'_K \cdot \overrightarrow{n}_K} \;=\; 0,$$

We rewrite this expression in terms of jumps between elements:

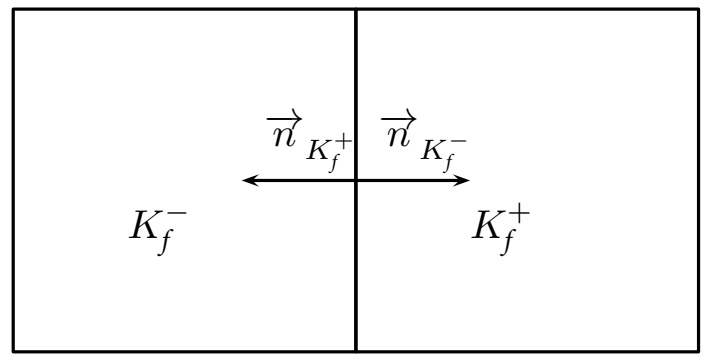
$$\sum_{F\in\mathcal{F}_{int}} \int_F \nabla p_F \cdot \overline{[\![p']\!]_F} - p_F \, \overline{[\![\nabla p']\!]_F} + \sum_{F\in\mathcal{F}_{ext}} \int_F \nabla p_F \cdot \overrightarrow{n}_{K_F} \, \overline{p'_{K_F}} - p_F \, \overline{\nabla p'_{K_F} \cdot \overrightarrow{n}_{K_F}} \;=\; 0,$$

assuming the solution is continuous on F

- p_F and $\overrightarrow{v}_F$ are the respective trace p and $\overrightarrow{v}$ sur F,
- $\overrightarrow{n}_F$ is the outgoing normal vector to the exterior face F.

The **jumps** are defined by

$$\begin{cases} [\![\overrightarrow{v}]\!]_F = \overrightarrow{v}_{K_F^+} \cdot \overrightarrow{n}_{K_F^+} + \overrightarrow{v}_{K_F^-} \cdot \overrightarrow{n}_{K_F^-}, \\ [\![p]\!]_F = p_{K_F^+} \overrightarrow{n}_{K_F^+} + p_{K_F^-} \overrightarrow{n}_{K_F^-}, \end{cases}$$



Numerical traces

Numerical traces are introduced to approximate ∇p_F and p_F. If they are stamped by the sign *hat, we consider the formula:*

$$\sum_{F\in\mathcal{F}_{int}} \int_F \widehat{\nabla p_F}\cdot\overline{[\![p']\!]_F} - \widehat{p_F}\ \overline{[\![\nabla p']\!]_F} + \sum_{F\in\mathcal{F}_{ext}} \int_F \widehat{\nabla p_F}\cdot\overrightarrow{n}_{K_F}\ \overline{p'_{K_F}} - \widehat{p_F}\ \overline{\nabla p'_{K_F}\cdot\overrightarrow{n}_{K_F}} \ = \ 0,$$

and for the exact solution, we have:

$$\widehat{\nabla p_F} \ = \ \nabla p_F \quad \textit{and} \quad \widehat{p_F} = p_F$$

The numerical traces are defined by:

$$\widehat{\nabla p_F} \ = \ \frac{\nabla p_{K+} + \nabla p_{K-}}{2} + \gamma[\![p]\!]_F + \delta[\![\nabla p]\!]_F \quad \textit{and} \quad \widehat{p_F} = \frac{p_{K+} + p_{K-}}{2} + \alpha\cdot[\![p]\!]_F + \beta[\![\nabla p]\!]_F$$

For the boundary condition, we have:

$$\widehat{\nabla p_F} \ = \ \nabla p_K + \beta'\Big(\nabla p_F\cdot n - ikp_F - \varphi\Big) \quad \textit{et} \quad \widehat{p_F} = p_K + \alpha'\Big(\nabla p_F\cdot n - ikp_F - \varphi\Big)$$

We get a generalization of the trace concept for discontinuous functions.

The choice of coefficients α, β, γ and δ

Objective: get coercive formulations adapted to iterative solvings

- Arnold's IPDG formulations (in classical DG framework)

$$\alpha = 0, \quad \beta = 0, \quad \gamma = \frac{\gamma_0}{h^2}, \quad \delta = 0$$

We end up with a coercive formulation but it doesn't go through an iterative process

- Outgoing flux method

$$\begin{aligned} &\text{ingoing flux :} && \nabla p \cdot n - ikp, \\ &\text{outgoing flux :} && \nabla p \cdot n + ikp. \end{aligned}$$

The numerical trace is a linear combination of the two outgoing traces:

$$\widehat{p_F} \quad = \quad \frac{1}{2ik}\left(\nabla p_{K_F^+} \cdot n_{K_F^+} + ikp_{K_F^+}\right) + \frac{1}{2ik}\left(\nabla p_{K_F^-} \cdot n_{K_F^-} + ikp_{K_F^-}\right).$$

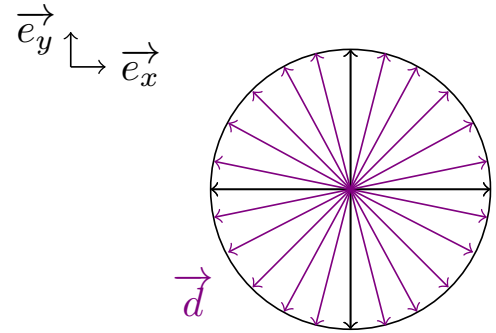
It corresponds to:

$$\alpha = 0 \text{ et } \beta = \frac{1}{2ik}.$$

Ditto for $\widehat{\nabla p_F}$.

Here, we get a coercive and iterative formulation.
This approach is difficult to generalize to complex equations.

The classical discrete Trefftz space

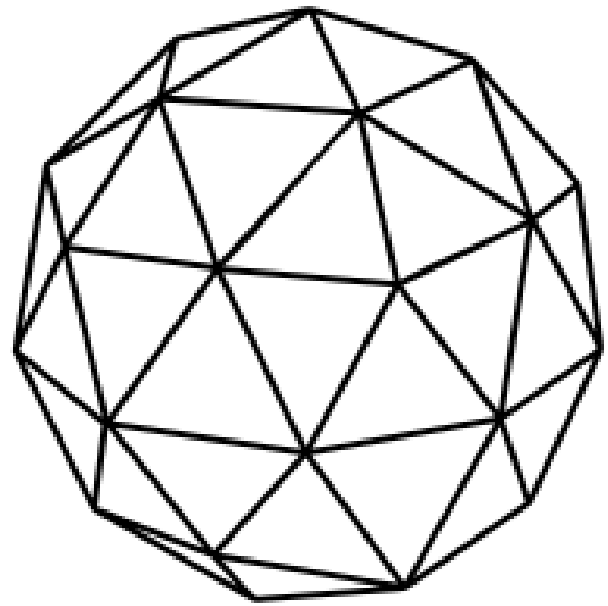


The basis functions of the discrete space are:

$$w_{d_n}(\mathbf{x}) \;=\; \exp\Big(ikd_n \cdot \mathbf{x}\Big),$$

with

- d_n, the direction of the plane wave
- $k = \frac{\omega}{c}$, the wave number



The basis functions are discontinuous:

$$p_h(\mathbf{x}) = \sum_{i=1}^{I} [p_h]_{K,i} w_{d_i}(\mathbf{x})(\mathbf{x})$$

This discretization leads to

- easy-to-perform analytical calculations
- a very ill-conditioned linear system.

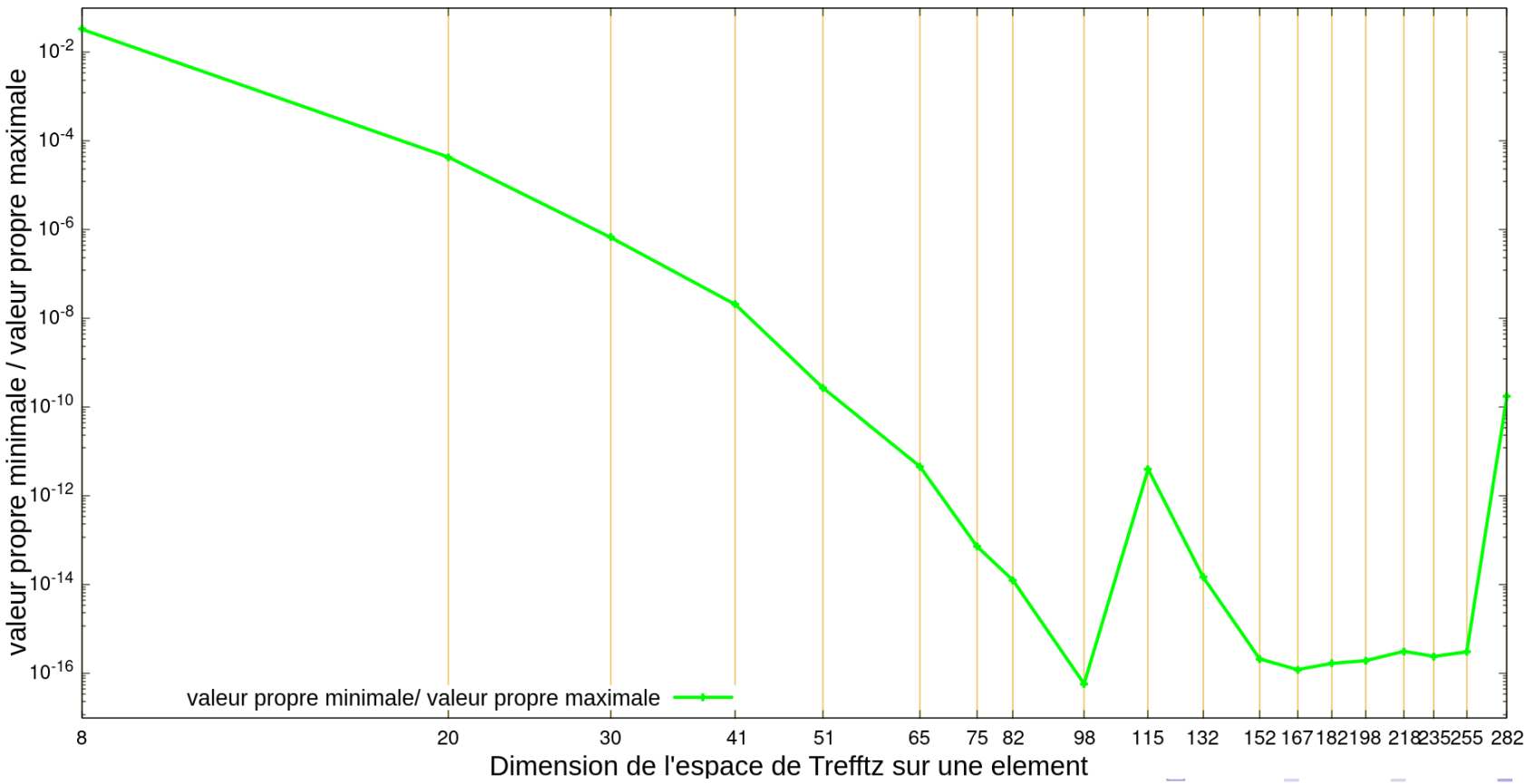
Conditioning problem

Theoretically, plane waves form a free family.

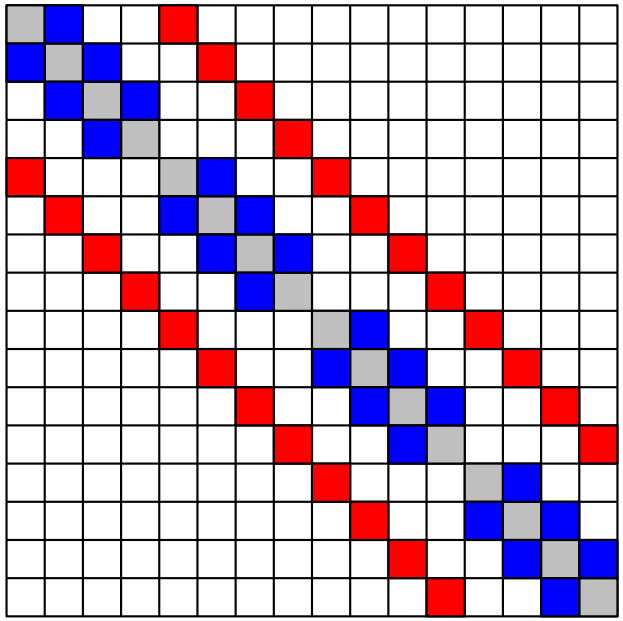
$$w_{d_n}(\mathbf{x}) \;=\; \exp\Big(ikd_n \cdot \mathbf{x}\Big),$$

Numerically, this family is almost linked.

$$R_{m,m'} \;=\; \int_K w_{d_m}(\mathbf{x})\,\overline{w_{d_{m'}}(\mathbf{x})}$$



The matrix skeleton



K_{13}	K_{14}	K_{15}	K_{16}
K_9	K_{10}	K_{11}	K_{12}
K_5	K_6	K_7	K_8
K_1	K_2	K_3	K_4

K_{13}	K_{14}	K_{15}	K_{16}
K_9	K_{10}	K_{11}	K_{12}
K_5	K_6	K_7	K_8
K_1	K_2	K_3	K_4

- the basis function of element 1 interact with those of elements 1, 2 and 5
- The basis functions of element 6 interact with those of elements 2, 5, 6, 7 and 10.

This very structured feature favors many optimizations:

- Full blocks adapted to parallelism OpenMp
- A block sparse structure adapted to massively parallel computing
- Assembling can be avoided

Reading

- O. Cessenat, and B. Després. "Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problem." SIAM journal on numerical analysis. 1998.
- P. Monk, and D. Q. Wang. A least-squares method for the Helmholtz equation. Computer methods in applied mechanics and engineering. 1999.
- A. Moiola. Trefftz-discontinuous Galerkin methods for time-harmonic wave problems. Diss. ETH Zurich. 2011.
- R. Hiptmair, A. Moiola, and I. Perugia. "A survey of Trefftz methods for the Helmholtz equation." Building bridges: connections and challenges in modern approaches to numerical partial differential equations. Springer, 2016.
- M. Sirdey, Méthode itérative de type Trefftz pour la simulation d'ondes électromagnétiques en trois dimensions, diss. UPPA, 2022

Section 3

Extension of Trefftz methods to general hyperbolic systems

Revisiting Trefftz methods with the hyperbolic formalism

Wave propagation problems are

- quasi-elliptic at the local level.

$$\Delta p + \omega^2 p = 0$$

- hyperbolic over long distances.

$$\begin{cases} -i\omega \mathbf{v} = \nabla p \\ -i\omega p = \operatorname{div}(\mathbf{v}) \end{cases} \iff \begin{cases} \frac{\partial \mathbf{v}}{\partial t} = \nabla p & \text{with } p(x,t) = p(x)\exp(-i\omega t), \\ \frac{\partial p}{\partial t} = \operatorname{div}(\mathbf{v}) & \text{with } \mathbf{v}(x,t) = \mathbf{v}(x)\exp(-i\omega t), \end{cases}$$

Is it possible to better understanding wave problems by introducing Friedrichs' formalism?

- what is a Friedrichs' system?
- What basic information does it contain?
- How to understand Riemann's solvers in this framework?
- What boundary conditions?

Friedrichs' systems

Definition of symmetrical Friedrichs systems

Find $Y : \Omega \times \mathbb{R}^+ \longrightarrow \mathbb{R}^m$ with $\Omega \subset \mathbb{R}^d$

$$M(\mathbf{x})\frac{\partial Y}{\partial t}(\mathbf{x},t) + \sum_{j=1}^{d} \frac{\partial}{\partial x_j} F_j(Y)(\mathbf{x},t) = 0$$

- $M : \Omega \longrightarrow \mathbb{R}^{m\times m}$ is a field of positive definite symmetric matrices
- The symbol for the differentiation operator in space

$$F : \mathbb{R}^d \longrightarrow \mathbb{R}^{m\times m} \quad \text{avec} \quad F[\xi] = \sum_{j=1}^{d} \xi_j \; F_j$$

with $F_j \in \mathbb{R}^{m\times m}$, $1 \leq j \leq d$, symmetrical matrices.

Theorem: Energy conservation formula

$$\frac{\partial}{\partial t}\int_\Omega \frac{Y^T(\mathbf{x},t)M(\mathbf{x})Y(\mathbf{x},t)}{2}d\mathbf{x} + \frac{1}{2}\int_{\partial\Omega} Y^T(\mathbf{x},t)F[\overrightarrow{n}]Y(\mathbf{x},t) \;=\; 0$$

- First term: total energy
- Second term: work of external forces

First example of Friedrichs system: acoustics in 3D

$$\left.\begin{array}{l}\rho\frac{\partial \mathbf{v}}{\partial t}(\mathbf{x},t)+\nabla p(\mathbf{x},t)=0,\\ \chi\frac{\partial p}{\partial t}(\mathbf{x},t)+\mathrm{div}\ \ \mathbf{v}(\mathbf{x},t)=0.\end{array}\right\}\Longleftrightarrow\left\{\begin{array}{l}\rho\frac{\partial \mathbf{v}_1}{\partial t}(\mathbf{x},t)+\frac{\partial p}{\partial x_1}(\mathbf{x},t)=0,\\ \rho\frac{\partial \mathbf{v}_2}{\partial t}(\mathbf{x},t)+\frac{\partial p}{\partial x_2}(\mathbf{x},t)=0,\\ \rho\frac{\partial \mathbf{v}_3}{\partial t}(\mathbf{x},t)+\frac{\partial p}{\partial x_3}(\mathbf{x},t)=0,\\ \chi\frac{\partial \rho}{\partial t}(\mathbf{x},t)+\frac{\partial \mathbf{v}_1}{\partial x_1}(\mathbf{x},t)+\frac{\partial \mathbf{v}_2}{\partial x_2}(\mathbf{x},t)+\frac{\partial \mathbf{v}_3}{\partial x_3}(\mathbf{x},t)=0.\end{array}\right.$$

The unknown and the hyperbilic matrices

$$Y=\begin{bmatrix}v_1\\v_2\\v_3\\p\end{bmatrix},\quad F_1=\begin{bmatrix}0&0&0&1\\0&0&0&0\\0&0&0&0\\1&0&0&0\end{bmatrix},\quad F_2=\begin{bmatrix}0&0&0&0\\0&0&0&1\\0&0&0&0\\0&1&0&0\end{bmatrix},\quad F_3=\begin{bmatrix}0&0&0&0\\0&0&0&0\\0&0&0&1\\0&0&1&0\end{bmatrix}$$

The hyperbolic symbol

$$M=\begin{bmatrix}\rho&0&0&0\\0&\rho&0&0\\0&0&\rho&0\\0&0&0&\chi\end{bmatrix}\qquad F[\xi]=\begin{bmatrix}0&0&0&\xi_1\\0&0&0&\xi_2\\0&0&0&\xi_3\\\xi_1&\xi_2&\xi_3&0\end{bmatrix}$$

- Total energy of the system: $\displaystyle\int_\Omega \frac{Y^TMY}{2}=\int_\Omega\frac{\rho|\mathbf{v}|^2+\chi p^2}{2}d\mathbf{x}$
- Work of external forces: $\displaystyle\int_{\partial\Omega}\frac{Y^TF[n]Y}{2}=\int_{\partial\Omega}p\mathbf{v}\cdot\mathbf{n}ds_{\mathbf{x}}$

Second example of Friedrichs' system: Maxwell system in 3D

$$\begin{cases} \varepsilon_0(\mathbf{x})\dfrac{\partial \mathbf{E}}{\partial t} - \operatorname{rot}\mathbf{H} = 0, \\ \\ \mu_0(\mathbf{x})\dfrac{\partial \mathbf{H}}{\partial t} + \operatorname{rot}\mathbf{E} = 0. \end{cases}$$

The unknown and the hyperbolic matrices

$$Y = \begin{bmatrix} E_1 \\ E_2 \\ E_3 \\ H_1 \\ H_2 \\ H_3 \end{bmatrix}, \quad F_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad F_2 = \cdots, \quad F_3 = \cdots$$

$$M = \begin{bmatrix} \varepsilon_0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \varepsilon_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \varepsilon_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu_0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu_0 \end{bmatrix} \quad F[\xi] = \begin{bmatrix} 0 & 0 & 0 & 0 & \xi_3 & -\xi_2 \\ 0 & 0 & 0 & -\xi_3 & 0 & \xi_1 \\ 0 & 0 & 0 & \xi_2 & -\xi_1 & 0 \\ 0 & -\xi_3 & \xi_2 & 0 & 0 & 0 \\ \xi_3 & 0 & -\xi_1 & 0 & 0 & 0 \\ -\xi_2 & \xi_1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- Total energy of the system: $\displaystyle\int_\Omega \frac{Y^T M Y}{2} = \int_\Omega \frac{\varepsilon_0|\mathbf{E}|^2 + \mu_0|\mathbf{H}|^2}{2} d\mathbf{x}$
- Work of external forces: $\displaystyle\int_{\partial\Omega} \frac{Y^T F[n] Y}{2} = -\int_{\partial\Omega} (\mathbf{E}\times\mathbf{H})\cdot\mathbf{n} ds_{\mathbf{x}}$

Is it really general ?

- Elastic wave equation
- Anisotropic acoustics
- Anisotropic elastic wave equation
- Anisotropic electromagnetism
- In some situations in Aeroacoustics

The matrix $\mathbf{F}_i$ sould be constant to have a Galerkin method.

The matrix M can vary !

If the matrix $\mathbf{F}_i$ are not symmetric, you should symmetrize them.

Bibliographie

- Symmetric Hyperbolic Linear Differential Equations, Friedrichs (1959)
- Hyperbolic Partial Differential Equations and Geometric Optics Rauch (2012)
- Monk, Richter, A Discontinuous Galerkin Method for Linear Symmetric Hyperbolic Systems in Inhomogeneous Media. J Sci Comput 22, 443–477 (2005)

Hyperbolic problems in frequency regime

We seek a solution with a harmonic time dependency:

$$Y(\mathbf{x},t) \;=\; \Re(Y(\mathbf{x})\;\exp(-i\omega t))$$

Find $Y \in L^2(\Omega)$ such that

$$\begin{cases} -i\omega M(\mathbf{x})Y(\mathbf{x}) + \displaystyle\sum_{i=1}^{n} \frac{\partial}{\partial x_i} F_i Y(\mathbf{x}) = 0, & \mathbf{x} \in \Omega, \\ F_-[n]Y(\mathbf{x}) = \varphi(\mathbf{x}), & \mathbf{x} \in \partial\Omega, \end{cases}$$

The variational space is:

$$H \;=\; \left\{ Y \in (L^2(\Omega))^k \text{ such that } \sum_{i=1}^{d} \frac{\partial F_i Y}{\partial x_i} \in (L^2(\Omega))^k \right\}$$

A very useful formula:

$$\int_\Omega \mathrm{div}\,(FY)\cdot\overline{Y'} \;+\; \int_\Omega Y\cdot\mathrm{div}\,(\overline{FY'}) \;=\; \int_{\partial\Omega} F[n]Y\cdot\overline{Y'} \qquad \text{pour tout } Y \text{ et } Y' \in H.$$

Riemann solver method in the heterogenous case

We start with

$$\sum_{K\in\mathcal{K}}\int_{\partial K} F[\mathbf{n}]Y\cdot\overline{Y'_K}=0$$

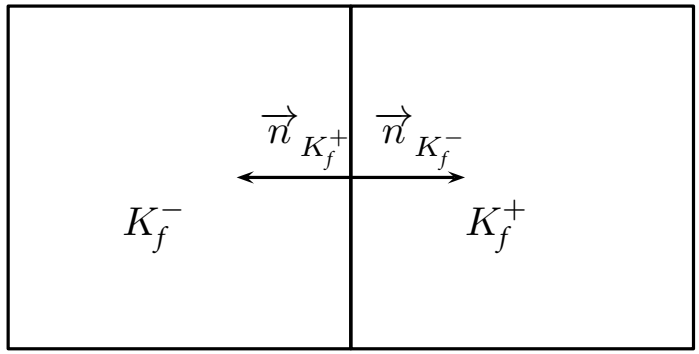
Then we focus on the faces and rewrite the previous formula as

$$\sum_{f\in\mathcal{F}_{int}}\int_f F[n_{K_F^+}]Y_f\cdot\overline{[\![Y']\!]_f}+\sum_{f\in\mathcal{F}_{ext}}\int_f F[\mathbf{n}]Y_f\cdot\overline{Y'_K}=0$$

with

- Y_f the trace of the exact solution
- the jumps defined by

$$[\![Y']\!]_f=Y'_{K_+^f}-Y'_{K_-^f}$$



The flux $F[n]Y_F$ is replaced by a numerical flux $\widehat{F[n]Y_f}$

$$\sum_{f\in\mathcal{F}_{int}}\int_f \widehat{F[n]Y_f}\cdot\overline{[\![Y']\!]_F}+\sum_{f\in\mathcal{F}_{ext}}\int_f \widehat{F[n]Y_f}\cdot\overline{Y'_K}=0$$

Then the question is: how to define the numerical flux $\widehat{F[n]Y_f}$?

Interior Riemann flux

- For $x_1 < 0$, we have

$$M_- \frac{\partial Y_-}{\partial t} + \sum_{i=1}^{d} \frac{\partial}{\partial x_i} F_i Y_- = 0.$$

- For $x_1 > 0$, we have

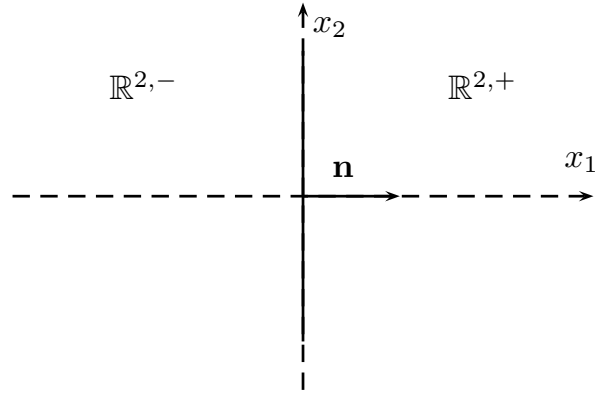
$$M_+ \frac{\partial Y_+}{\partial t} + \sum_{i=1}^{d} \frac{\partial}{\partial x_i} F_i Y_+ = 0.$$

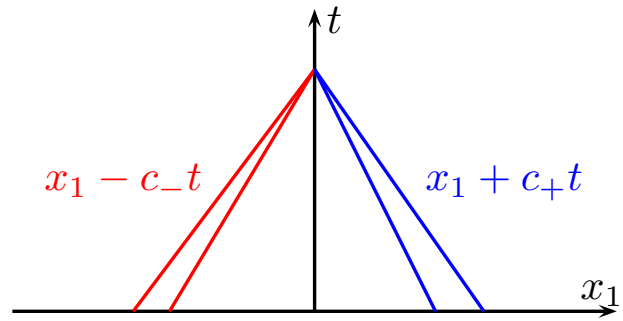
- Initial c $(t = 0)$

$$\begin{cases} Y(\mathbf{x}, 0) = Y_- & \text{pour } x_1 < 0, \\ Y(\mathbf{x}, 0) = Y_+ & \text{pour } x_1 > 0. \end{cases}$$

- Transmission conditions $(x_1 = 0)$.

$$F[e_1]Y_- \;=\; F[e_1]Y_+.$$





Condition initiale en $t = 0$

Apply the characteristics method to get:

$$Y(x_1, t) \text{ pour } x_1 \neq 0 \text{ et } t > 0 \quad \Longrightarrow \quad \widehat{F[n]Y}(0, t) = F[n]Y(0, t) \text{ pour } t > 0.$$

Interior Riemann flux

Solve the acoustic problem set in $\mathbb{R}^d$

- For $x_1 < 0$, we have

$$\begin{cases} A_- \dfrac{\partial \overrightarrow{v}}{\partial t}(\mathbf{x},t) + \nabla p(\mathbf{x},t) &= 0, \\ \chi_- \dfrac{\partial p}{\partial t}(\mathbf{x},t) + \operatorname{div} \overrightarrow{v}(\mathbf{x},t) &= 0, \end{cases}$$

- For $x_1 > 0$, we have

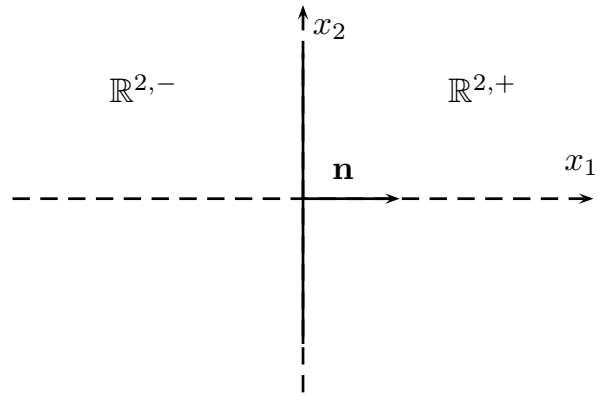
$$\begin{cases} A_+ \dfrac{\partial \overrightarrow{v}}{\partial t}(\mathbf{x},t) + \nabla p(\mathbf{x},,t) &= 0, \\ \chi_+ \dfrac{\partial p}{\partial t}(\mathbf{x},t) + \operatorname{div} \overrightarrow{v}(\mathbf{x},t) &= 0, \end{cases}$$

with initial conditions

$$\begin{cases} p(\mathbf{x},0) = p_- & \text{pour } x_1 < 0, \\ p(\mathbf{x},0) = p_+ & \text{pour } x_1 > 0, \\ \overrightarrow{v}(\mathbf{x},0) = \overrightarrow{v}_- & \text{pour } x_1 < 0, \\ \overrightarrow{v}(\mathbf{x},0) = \overrightarrow{v}_+ & \text{pour } x_1 > 0. \end{cases}$$

and the transmission conditions in $x_1 = 0$:

$$[\![p]\!]_F = 0 \quad \text{and} \quad [\![\overrightarrow{v}]\!]_F = 0$$



We then obtain

$$\widehat{p}(0,t) = \frac{Y_- p_- + Y_+ p_+}{Y_- + Y_+} + \frac{Y_- Y_+}{Y_- + Y_+} [\![\overrightarrow{v}]\!]_F,$$

$$\widehat{v}(0,t) = \frac{Z_- v_- + Z_+ v_+}{Z_- + Z_+} + \frac{Z_- Z_+}{Z_- + Z_+} [\![p]\!]_F$$

with admittance Y and impendance Z:

$$Y_\pm = \frac{1}{Z\pm} = \sqrt{\frac{n \cdot A_\pm n}{\chi_\pm}}$$

Conclusion for the section

- We have shown how to get the interior Riemann flux

$$\widehat{F[n]Y} = F[n]\frac{Y_{K_F^+} + Y_{K_F^-}}{2} + T[\![Y]\!]_f$$

- It is possible to do the same with the boundnary condition

$$\widehat{F[n]Y} = F[n]Y - \Big(F[n]_- Y - \varphi\Big) = F[n]_+ Y + \varphi$$

We end up with a Trefftz problem

$$\sum_{f\in\mathcal{F}_{int}} \int_f \widehat{F[n]Y_f} \cdot \overline{[\![Y']\!]_F} + \sum_{f\in\mathcal{F}_{ext}} \int_f \widehat{F[n]Y_f} \cdot \overline{Y'_K} = 0$$

This formulation is coercive !

Section 4

Improvement of discrete spaces

The problem of round-off errors

In 3D, the theory tells us that Trefftz methods are

- convergent
- well-posed

In practice

- Occasionally non-invertible
- Often non-convergent

Why?

- Two basis functions related to parallel directions are numerically linked.

Our objective: reduce the number of basis functionsby eliminating those that are unnecessary

- Decrease the computational costs
- Reduce the round-off errors

Two approaches:

- Basis reduction by principal component analysis (SVD)
- Quasi-Trefftz method

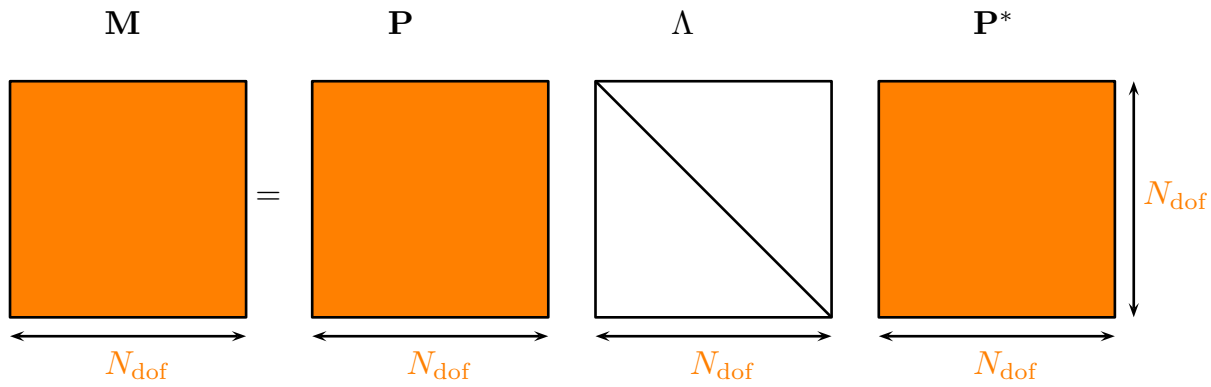
Basis reduction strategy

$\mathbf{M}$ is symmetrical positive definite

$$\mathbf{M}^{\mathrm{i,j}}_{\ell,k} = (\mathbf{w}^{\mathrm{i}}_{k}, \mathbf{w}^{\mathrm{i}}_{\ell})_{L^2_t(\partial\mathcal{T})}\, \delta_{\mathrm{i,j}}$$

pour $\mathrm{i,j} = 1, N_{\mathrm{elem}}$

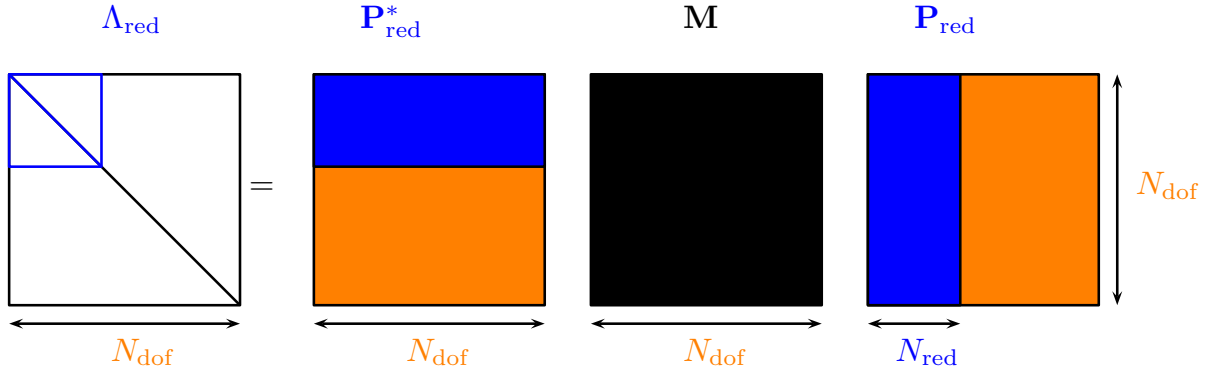
et $\quad \ell, k = 1, N.$



$\mathbf{P}$ is the orthogonal matrix formed with the eigenvectors.
Λ is the matrix of eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_{\mathrm{dof}}} \quad \text{avec } \lambda_{\mathrm{i}} = \Lambda_{\mathrm{i,i}}.$$

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{N_{\mathrm{red}}} \\ \vdots \\ \lambda_{N_{\mathrm{dof}}-1} \\ \lambda_{N_{\mathrm{dof}}} \end{pmatrix} \rightarrow \textbf{Filtering} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{N_{\mathrm{red}}} \end{pmatrix} \qquad \frac{\lambda_{\mathrm{i}}}{\lambda_1} < \varepsilon$$



Normalized reduced basis

The discrete space is reduced:

$$\mathbf{x} = \sum_{i=1}^{N} [\mathbf{x}]_i \mathbf{w}^i = \sum_{i=1}^{N} [\mathbf{y}]_i \widetilde{\mathbf{w}}^i$$

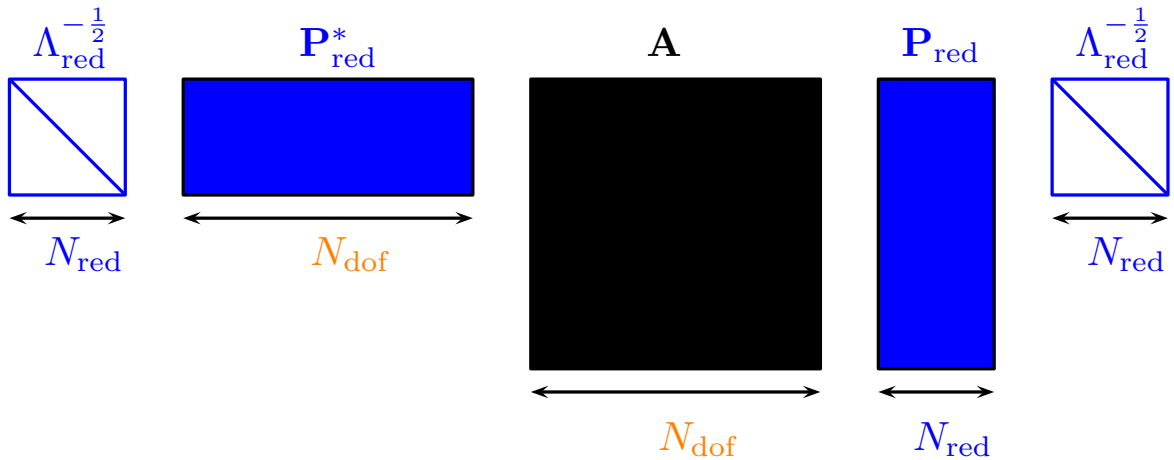
à

$$\mathbf{x}_{red} = \sum_{i=1}^{N_{red}} [\mathbf{y}]_i \widetilde{\mathbf{w}}^i, \quad \mathbf{Y}_{red} = [\mathbf{y}]_{i=1,N_{red}}.$$

Little information loss for $\frac{\lambda_i}{\lambda_1} < \varepsilon$, with ε small.

The reduced basis can be normalized and the new matrix is

$$\mathbf{A}_{red} = \mathbf{I}_{red} - \mathbf{N}_{red} =$$



The reduction of the basis improves the memory

$\mathbf{N}_{\mathrm{red}}$ depends on the size of element K and on the truncation criterion ε.

Size of K \ ε	10^{-13}	10^{-11}	10^{-9}	10^{-7}	10^{-5}	10^{-4}	10^{-3}	10^{-2}
0.25λ	154	126	96	70	48	48	30	24
0.5λ	190	186	174	132	96	84	70	48
1λ	196	196	196	190	180	174	148	114

0.25λ case, $N_{\mathrm{kry}}^{\mathrm{restart}} = 500$.

10^{-15}	ε	10^{-13}	10^{-11}	10^{-9}	10^{-7}	10^{-5}	10^{-4}	10^{-3}	10^{-2}
N = 196	N_{red}	154	126	96	70	48	36	30	16
1.57 Go	Memory cost (Go)	1.23	1.01	0.76	0.56	0.38	0.28	0.24	0.13

The iterative solution does not converge without any reduction

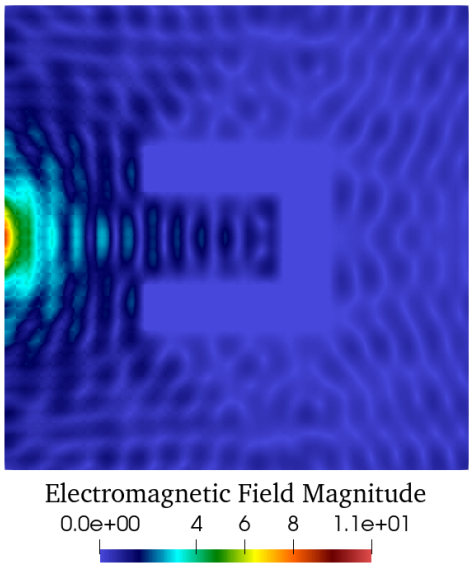
$\Longrightarrow$ Improvement of spectral properties

How far can we reduce the basis?

ε	10^{-7}	10^{-5}	10^{-4}	10^{-3}	10^{-2}
Erreur	N/A	3.4×10^{-2}	9.8×10^{-2}	0.15	0.74
Coût mémoire (Go)	0.56	0.38	0.28	0.24	0.13
Temps (heures)	55	27	11.3	7.23	0.58

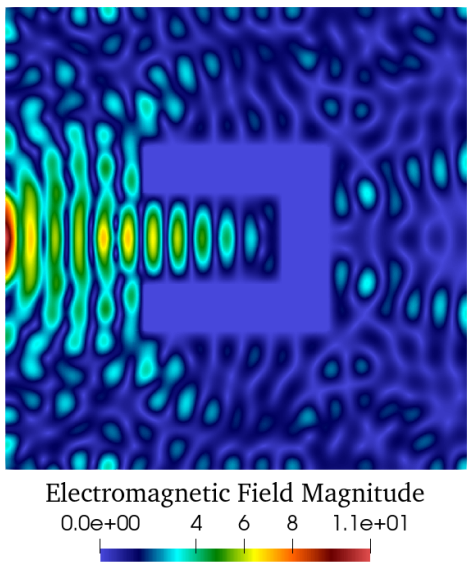
Too much reduction

$N_{\text{red}} = 16$ $\varepsilon = 10^{-2}$



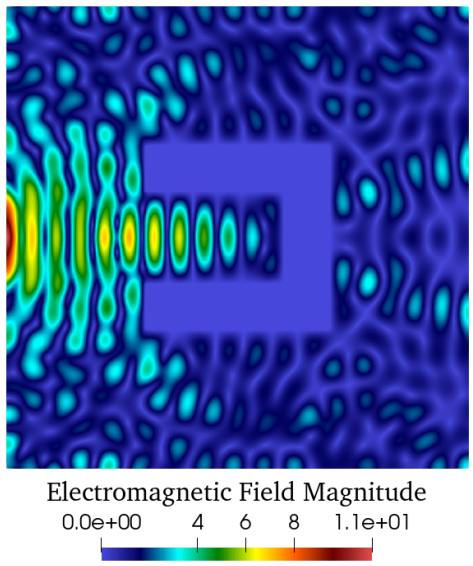
Good compromise

$N_{\text{red}} = 36$ $\varepsilon = 10^{-4}$



Reference solution

$N_{\text{red}} = 70$ $\varepsilon = 10^{-7}$



Quasi-Trefftz methods

- We introduce the solution operator

$$S_K : L^2(\partial K) \longrightarrow H^1(K) \quad \varphi \longmapsto p_K$$

$$\begin{cases} \Delta p_K + k^2 p_K = 0 & \text{dans } K, \\ \nabla p_K \cdot n - ikp_K = \varphi & \text{sur } \partial K, \end{cases}$$

  along with its approximation $S_{K,h} : \varphi \longrightarrow p_{K,h}$ for instance
  - finite elements
  - boundary finite elements
  - spectral differences and FR (Matthias Rivet, Sébastien Pernet)
  - quasi-analytical methods (Andréa Lagardère, LMIG, Guillaume Sylvand)
- A finite element subspace V_H of $L^2(\partial K)$ as well as the discrete Trefftz and quasi-Trefftz spaces

$$X_H = S_K V_H \text{ et } X_{H,h} = S_{K,h} V_H$$

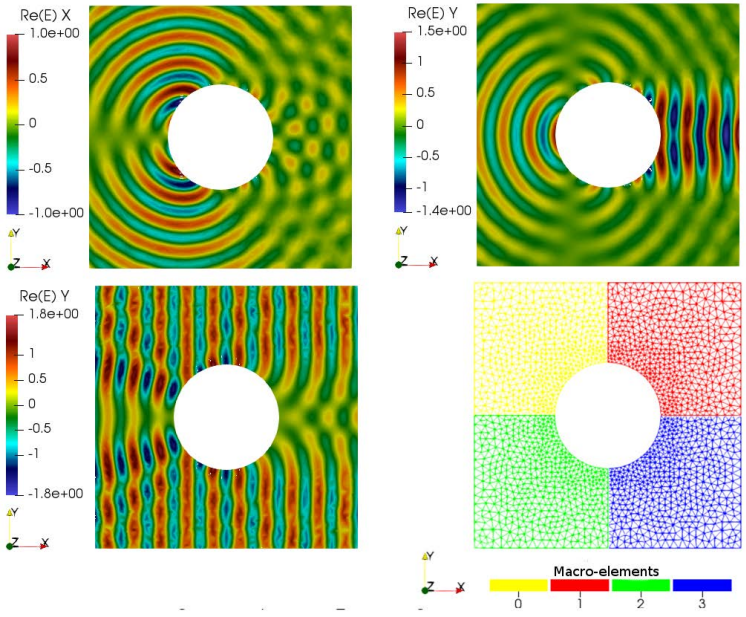
- A master Trefftz solver will couple the different solution methods (Previous formulation)

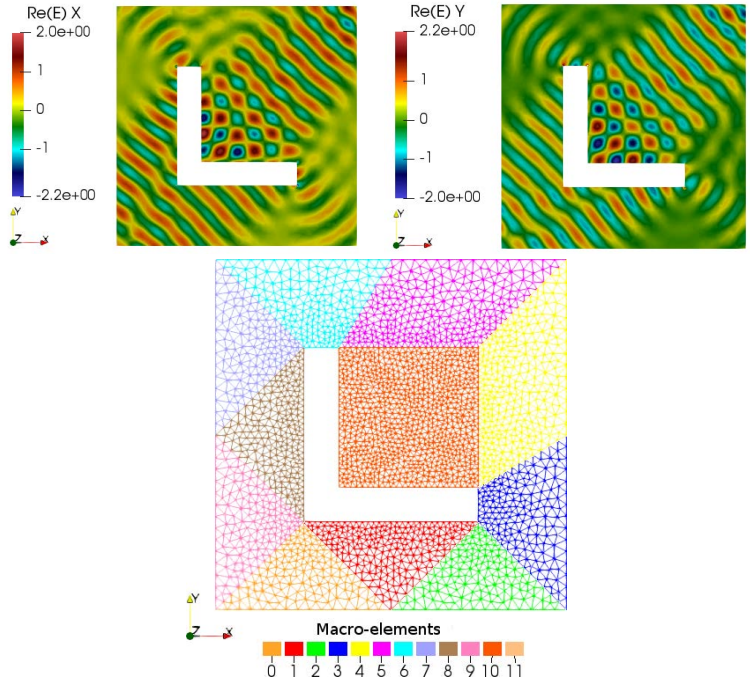
Computational cost

- A single LU factorization LU in one element and several solutions.
- Adapted to MPI.

  Local solution cost ($C\ N_{elem}$) $<<$ cost of the master solver ($C'\ N_{elem}^{\frac{4}{3}}$)

Two qualitative examples



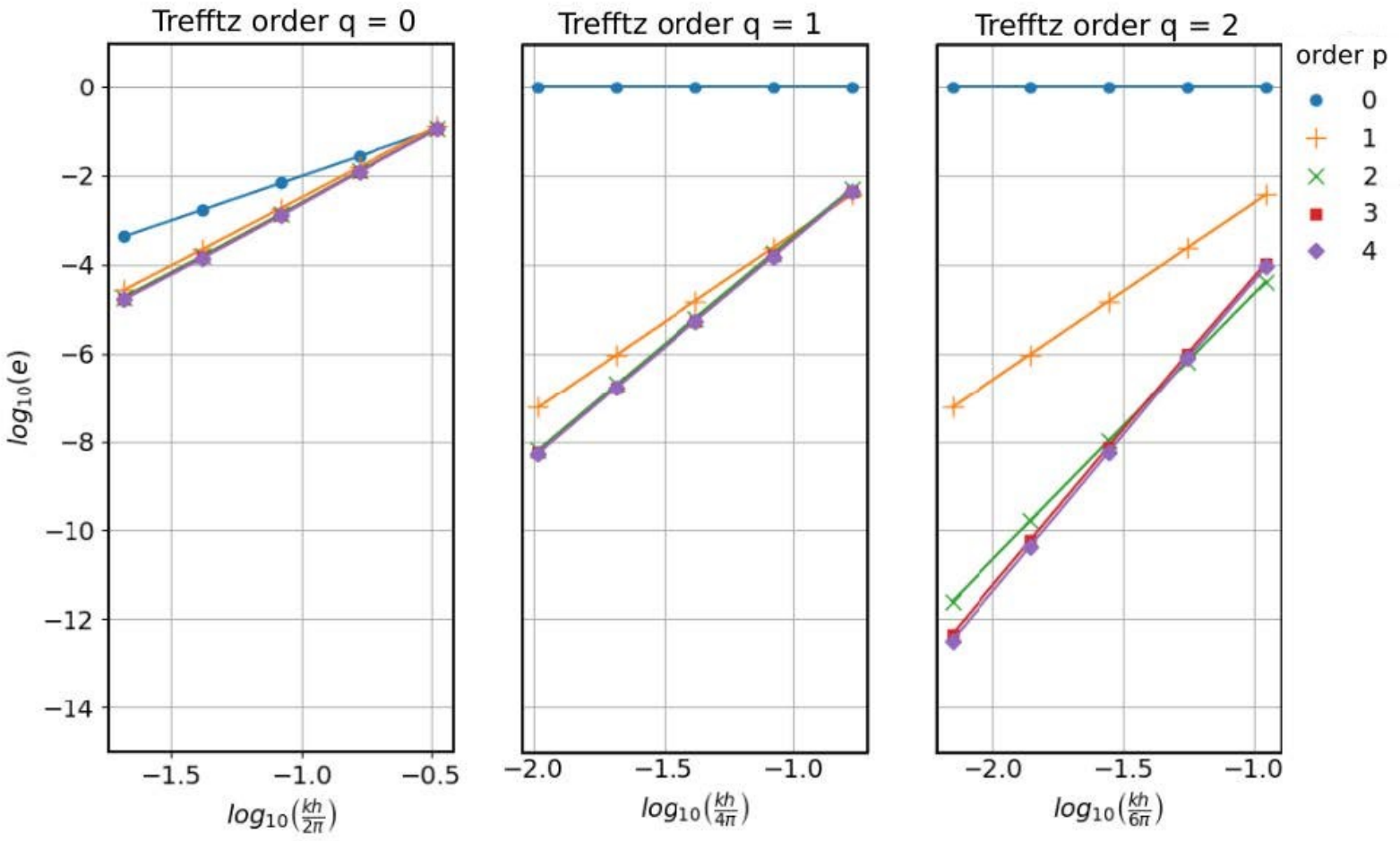


- Barucq, Bendali, Fares, Mattesi, Tordeux, A symmetric Trefftz-DG formulation based on a local boundary element method for the solution of the Helmholtz equation, JCP, (2017)
- Fure, Pernet, Sirdey, Tordeux, A discontinuous Galerkin Trefftz type method for solving the two dimensional Maxwell equations, PDE and Applications (2020)

Optimal convergence order

The order of the method is defined in two ways:

- the space order of the Trefftz method characterized by the trace space
- the order of the auxiliary method



Superconvergence of quasi-Trefftz methods

Expected order $+\frac{1}{2}$

Section 5

Implementation of Trefftz methods in a HPC environment

Structure of the matrix with a hexahedrical mesh

$$\mathbf{A}[\mathbf{x}] \;=\; \mathbf{F},$$

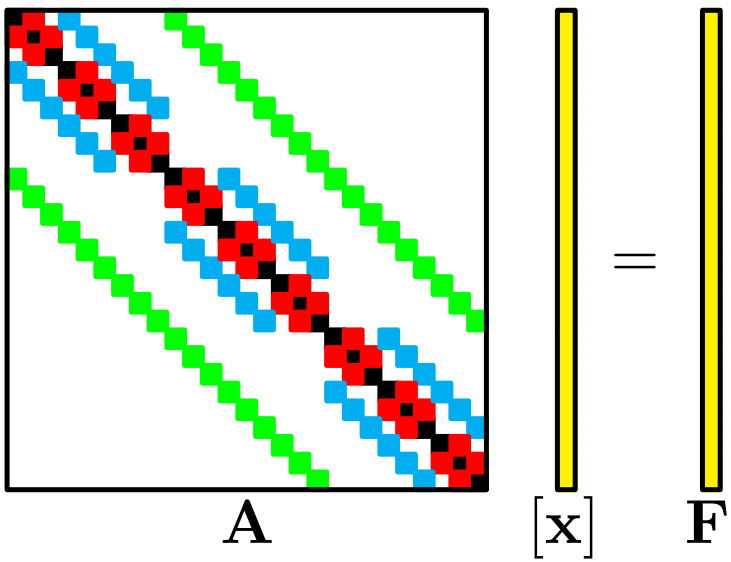
où

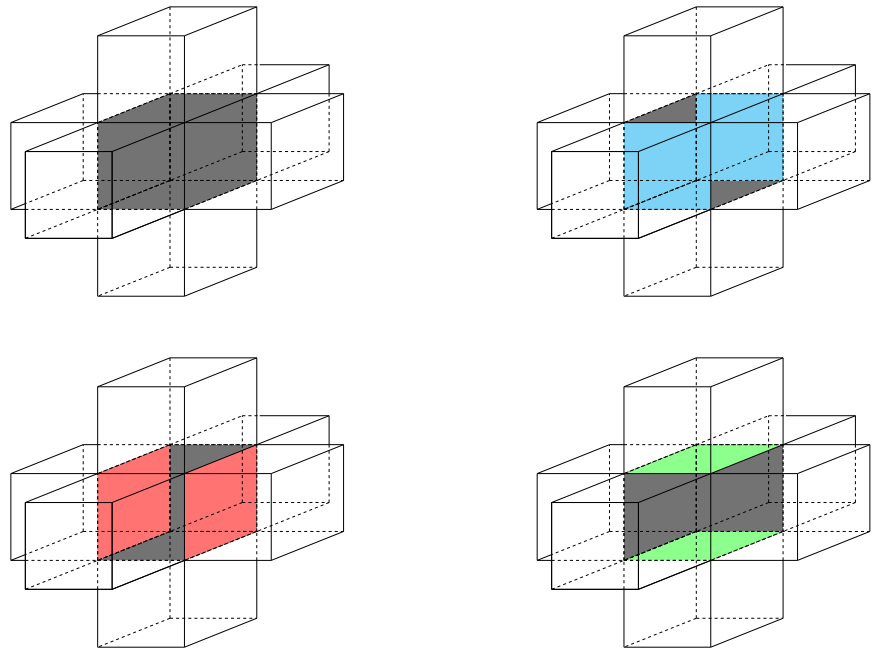
$$\mathbf{A}_{\mathrm{i,j}} := \mathbf{a}(\mathbf{w}^{\mathrm{j}}, \mathbf{w}^{\mathrm{i}})$$

$$\mathbf{F}_{\mathrm{i}} := \mathbf{l}(\mathbf{w}^{\mathrm{i}}) \quad \text{for} \;\; \mathrm{i,j} = 1, N_{\mathrm{ddl}}.$$

$$\mathit{taille}(\mathbf{A}) = N \times N_{\mathrm{elem}}$$

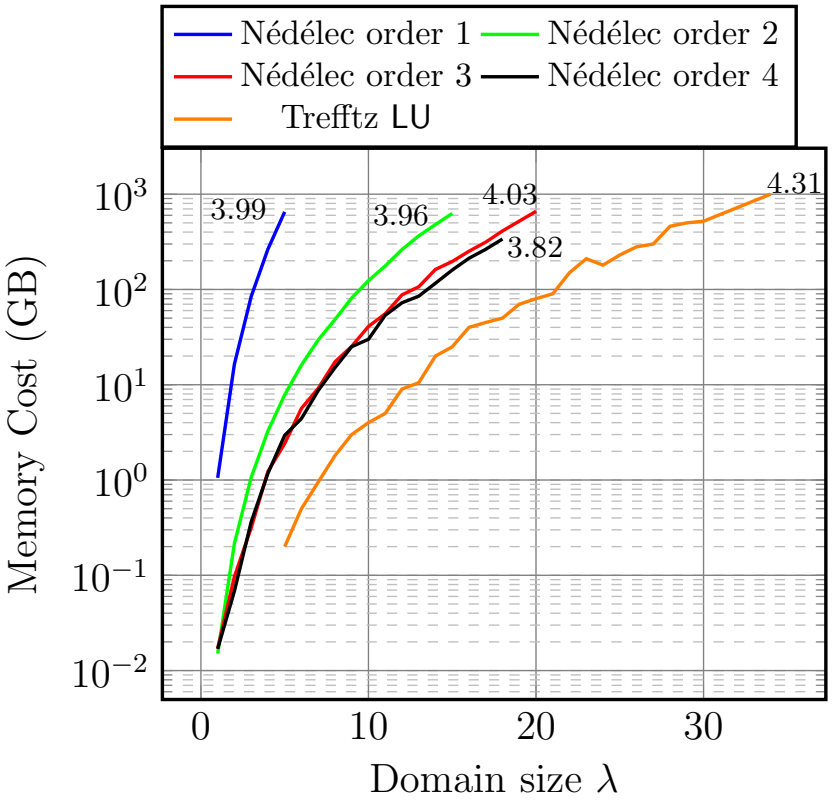
$$\mathit{nnz}(\mathbf{A}) = N^2 \times 7 \times N_{\mathrm{elem}}$$

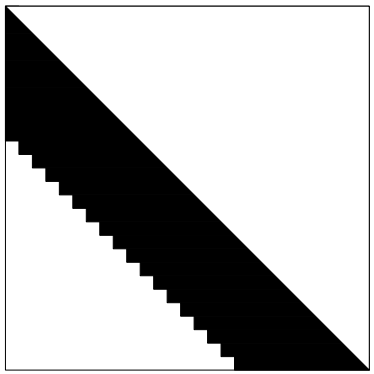




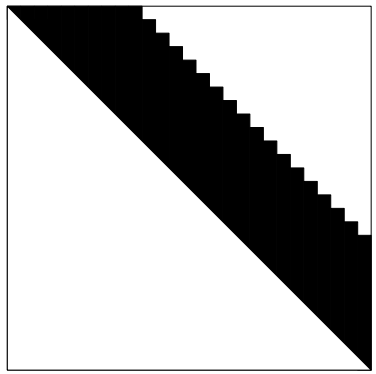
assembling of $\mathbf{A}$ for $N_{\mathrm{elem}} = 27 = 3^3$ cubes.

Memory limit of the direct solver





L

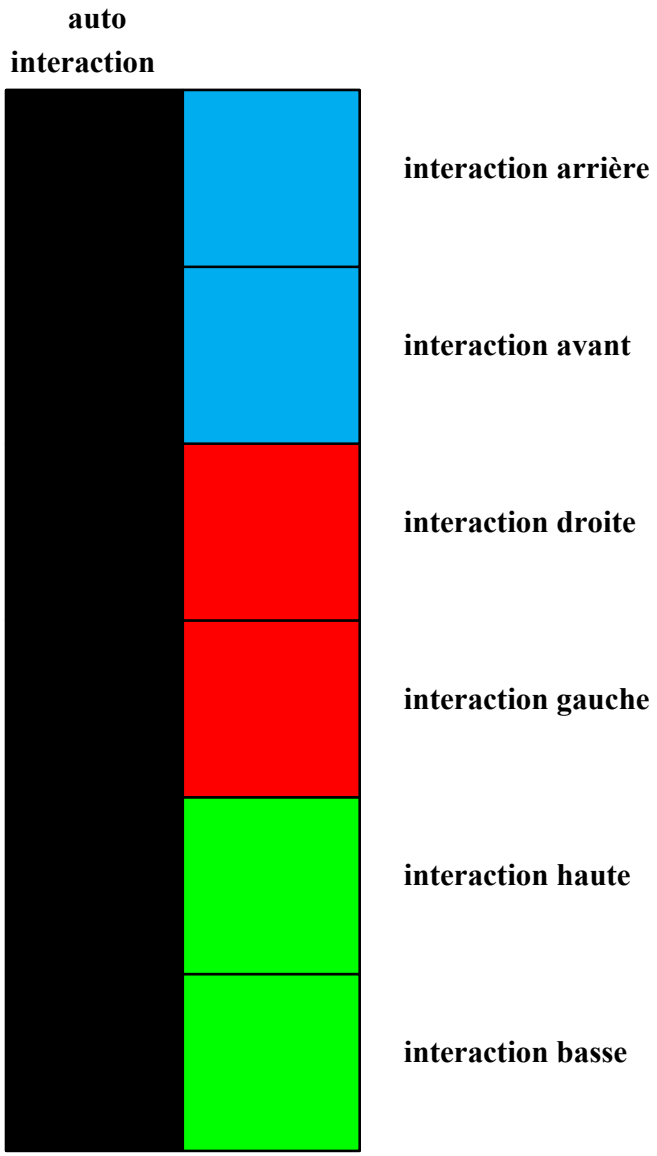


U

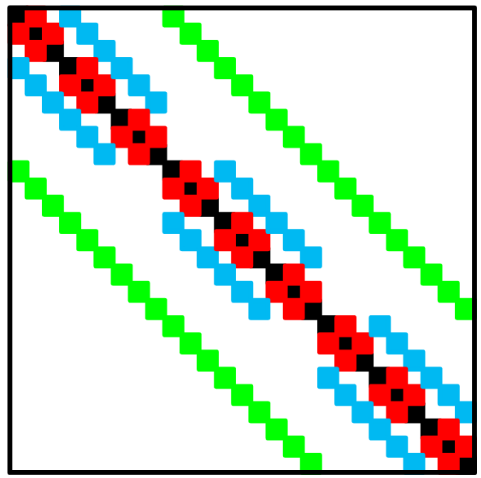
Domain size	5λ	10λ	20λ	100λ	200λ
Element number N_{elem}	512	1000	8000	10^6	8×10^6
Memory of LU	268 Mo	4.3 Go	69 Go	43 To	688 To

Table: Memory cost of LU factorization as a function of the domain size

Memory cost without assembly $7 \times N^2$



Cartesian mesh $\Longrightarrow$ structured matrix



$\mathbf{A}$

Without assembling, $\mathbf{Ax}$ can be computed for any $\mathbf{x}$

Memory cost without matrix assembly

- $N_{\text{kry}}^{restart} = 50$ the dimension of the Krylov space
- N_{elem} the element number
- $N = 52$ the number of plane waves
- $N_{\text{ddl}} = N \times N_{\text{elem}}$ the number of degrees of freedom

Memory cost with assembly and without assembly

$$\text{taille(Krylov)} + \text{nnz(A)} = 50 \times N \times N_{\text{elem}} + N^2 \times 7 \times N_{\text{elem}}$$

$$\text{size(Krylov)} + \text{size(interactions)} = 50 \times N \times N_{\text{elem}} + N^2 \times 7$$

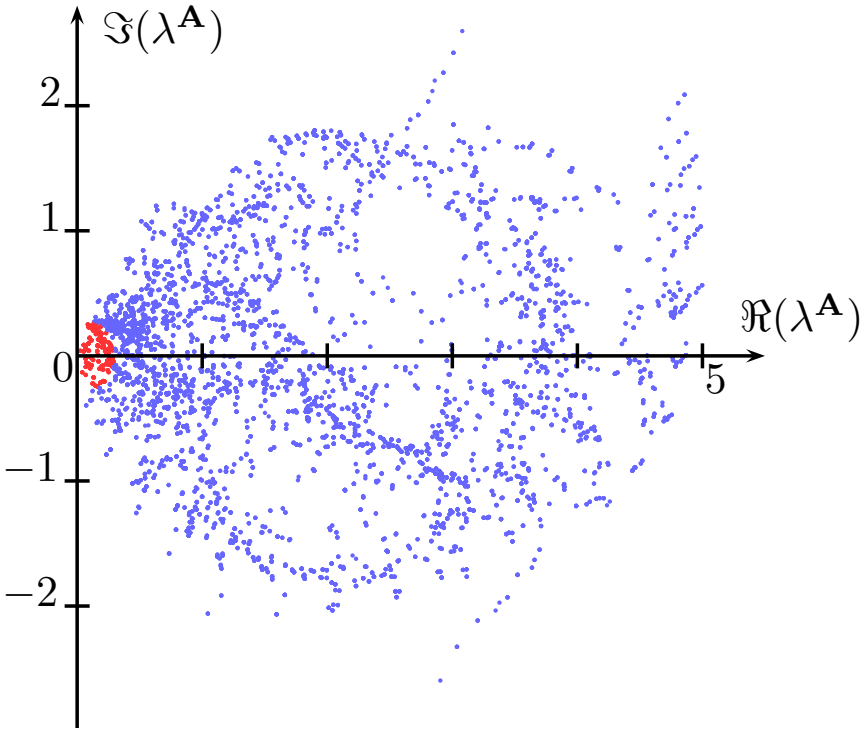
Without assembling: only the interactions are stored

Domain size	5λ	10λ	20λ	100λ	200λ
N_{elem}	512	1000	8000	10^6	8×10^6
with matrix assembly	180 Mo	344 Mo	2.75 Go	344 Go	2.75 To
Without matrix assembly	21.6 Mo	41.9 Mo	0.33 Go	41.6 Go	332 Go
$\frac{\text{taille(interactions)}}{\text{taille(Krylov)}}$	0.01	7×10^{-3}	9×10^{-4}	7×10^{-6}	9×10^{-7}

Cessenat-Després preconditioner

The convergence of Krylov method depends on
$\mathbf{A}$ **spectrum** and the distance between **eignevalues and** 0.

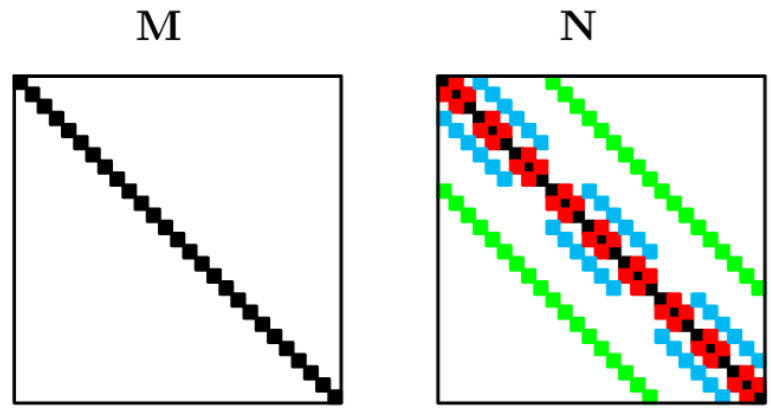
Idea: préconditionning $\mathbf{A}^{\#}\mathbf{A}[\mathbf{x}] = \mathbf{A}^{\#}\mathbf{F}$.



Real and imaginary parts of $\mathbf{A}$ spectrum, for a domain with size 6λ.

$$\mathbf{A} = \mathbf{M} + \mathbf{N}$$

Cessenat-Després decomposition [a]

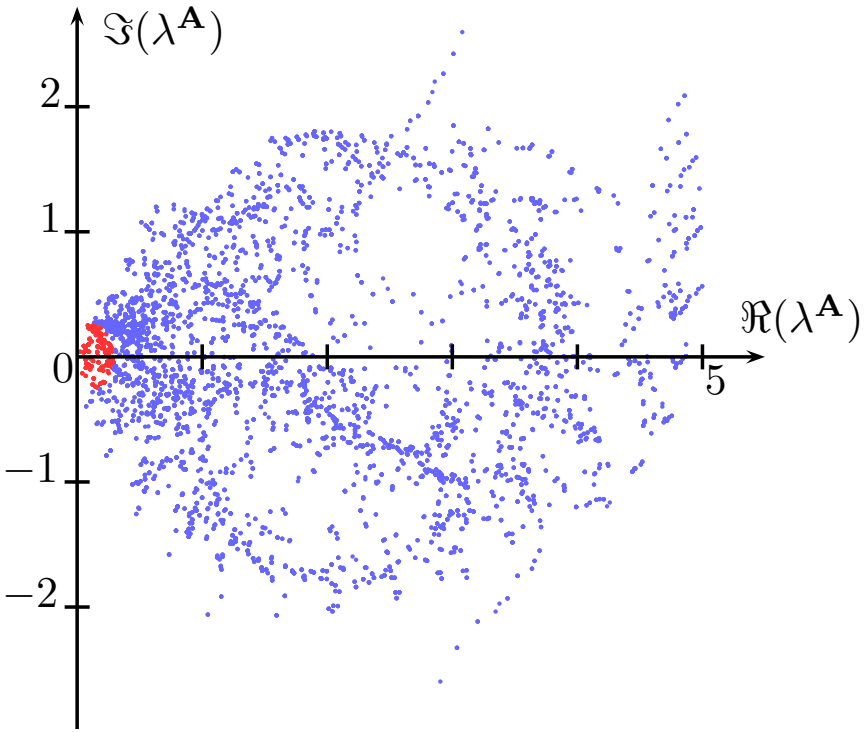


[a] O. Cessenat and B. Despres, (1998). Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problem. SIAM journal on numerical analysis, 35(1), 255-299.

Cessenat-Després preconditioner

The convergence of Krylov method depends on
A spectrum and the distance between **eignevalues and** 0.

Idée: $\mathbf{A}^{\#}\mathbf{A}[\mathbf{x}] = \mathbf{A}^{\#}\mathbf{F}$.



Real and imaginary parts of $\mathbf{A}$ spectrum, for a domain with size 6λ.

$$\begin{aligned} \mathbf{A}^{\#}(\mathbf{M}+\mathbf{N})[\mathbf{x}] &= \mathbf{A}^{\#}\mathbf{F} \\ \mathbf{A}^{\#}\mathbf{M}[\mathbf{x}]^{n+1} &= \mathbf{A}^{\#}\mathbf{N}[\mathbf{x}]^{n} + \mathbf{A}^{\#}\mathbf{F} \end{aligned}$$

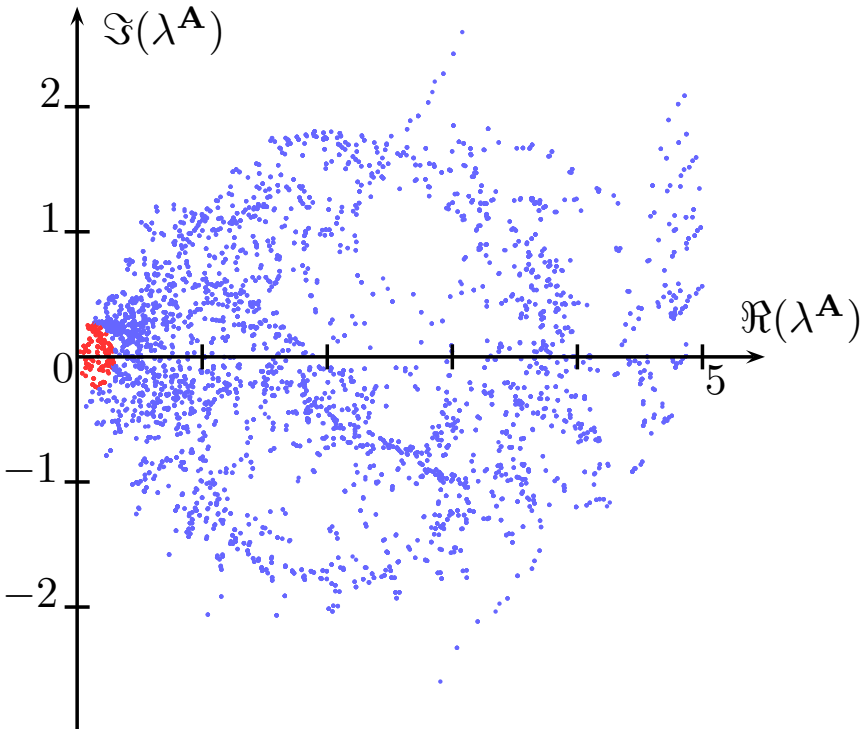
$$\rightarrow \mathbf{A}^{\#} = \mathbf{M}^{-1}$$

$$\widetilde{\mathbf{A}} = \mathbf{M}^{-1}\mathbf{A} = \mathbf{Id} + \mathbf{M}^{-1}\mathbf{N}$$

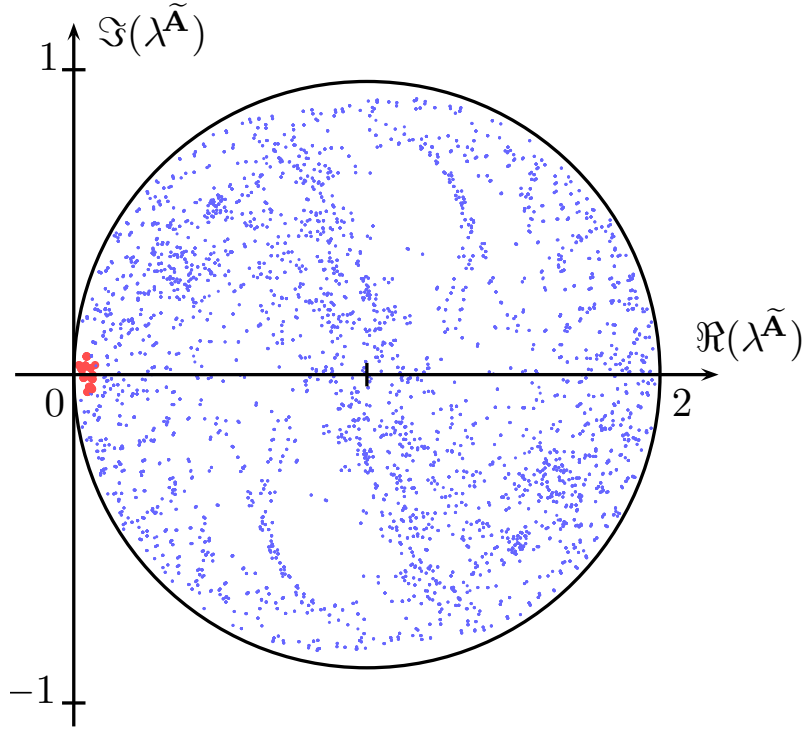
Cessenat-Després preconditioner

The convergence of Krylov method depends on
$\mathbf{A}$ spectrum and the distance between **eignevalues and** 0.

$\tilde{\mathbf{A}} = \mathbf{Id} + \mathbf{M}^{-1}\mathbf{N}$ est contractant



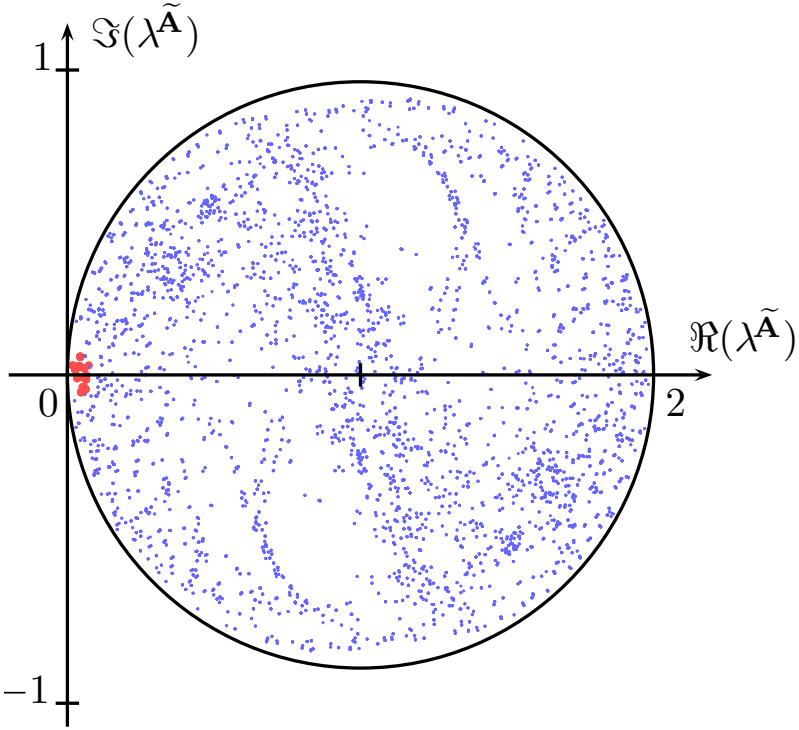
Real and imaginary parts of $\mathbf{A}$ spectrum, for a domain with size 6λ.



Real and imaginary part of $\widetilde{\mathbf{A}}$ spectrum, for a domain size 6λ.

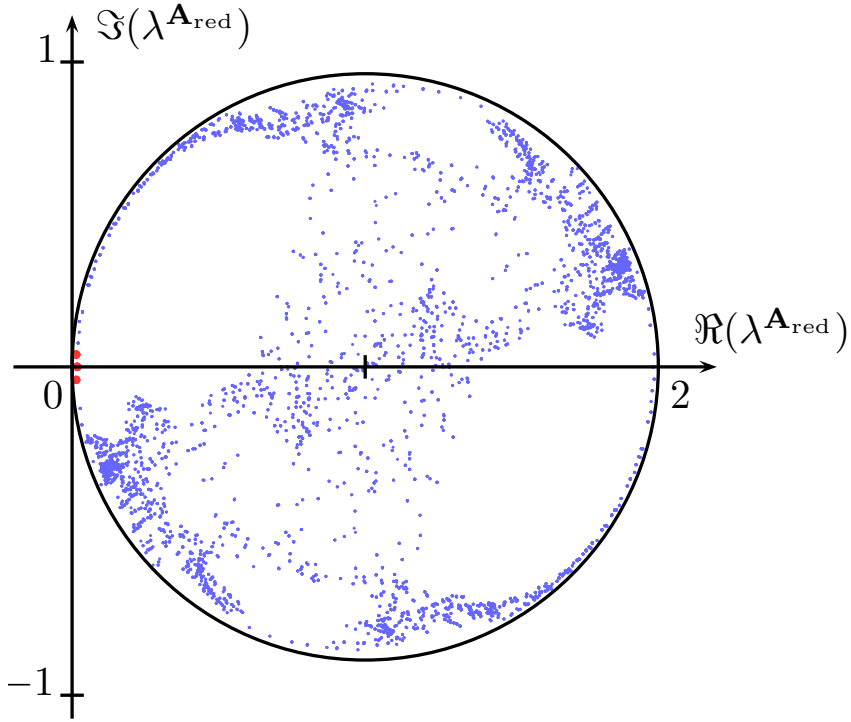
Basis reduction improves matrix conditioning

Preconditioned reduced matrix $\tilde{\mathbf{A}}$



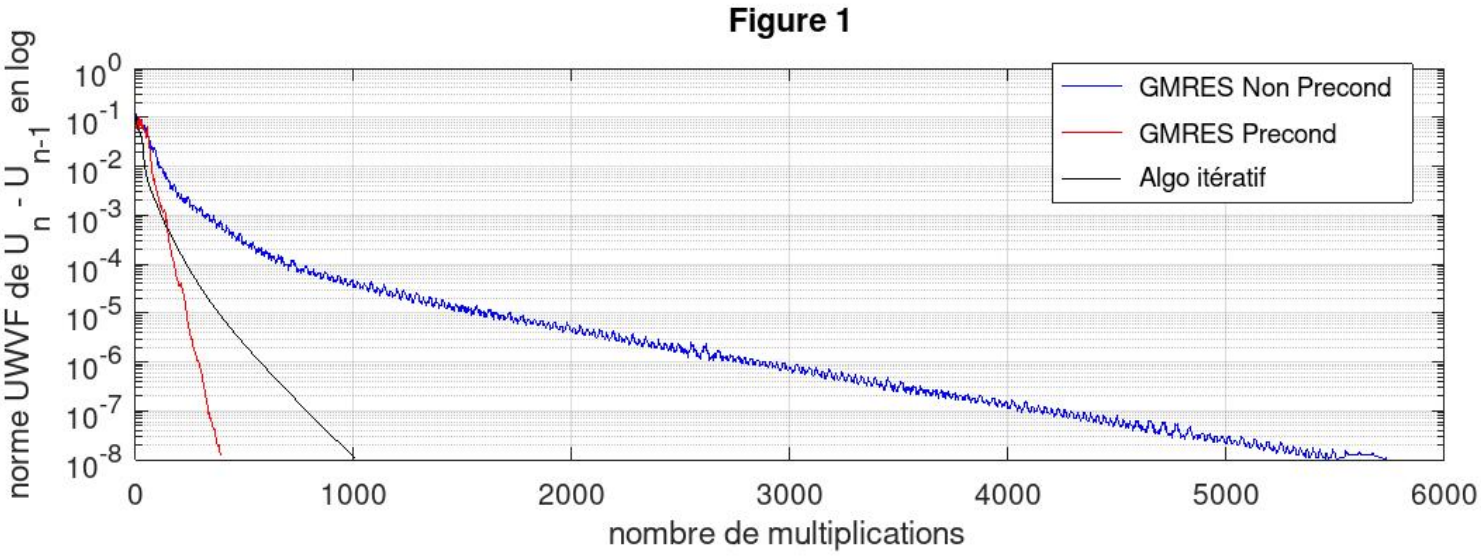
Real and imaginary parts of $\tilde{\mathbf{A}}$ spectrum, for a size domain 6λ.

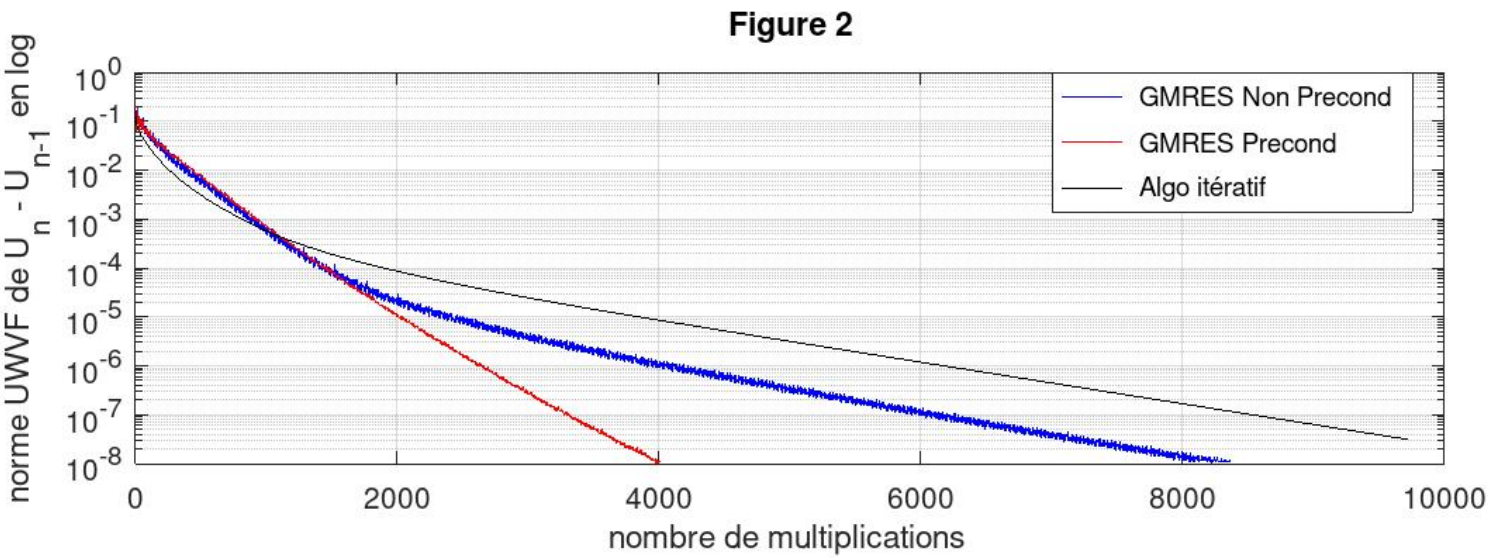
Reduced matrix $\mathbf{A}_{\text{red}}$



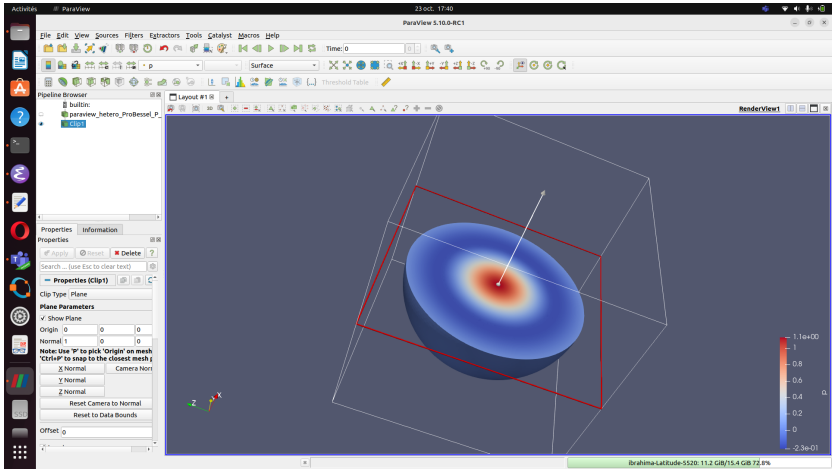
Real and imaginary parts of $\mathbf{A}_{red}$ spectrum, for a size domain 6λ.

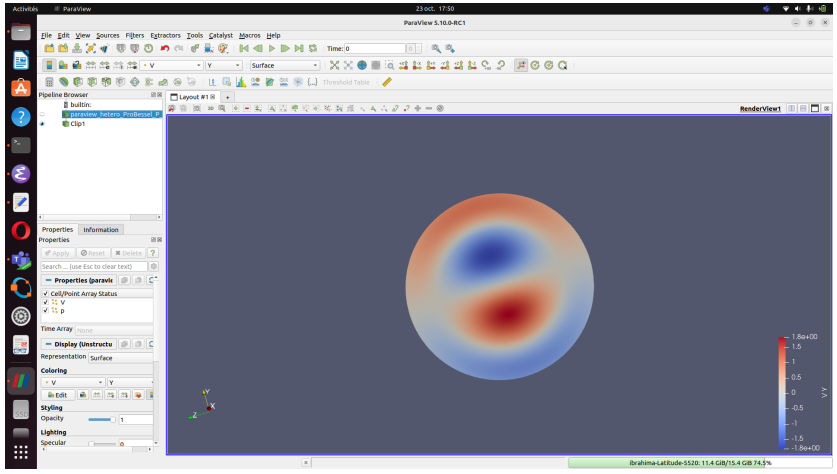
Iterative resolution of the linear system $AU = F$ using Krylov method

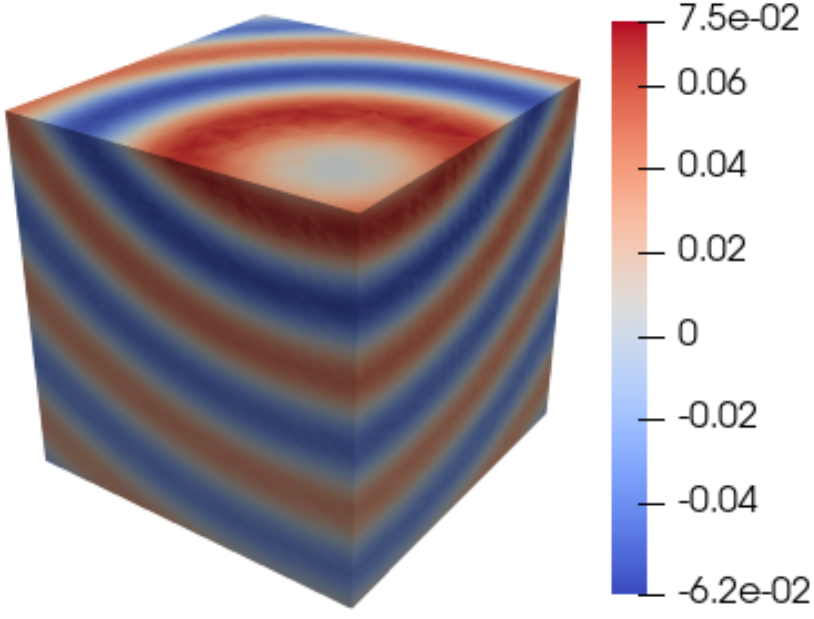


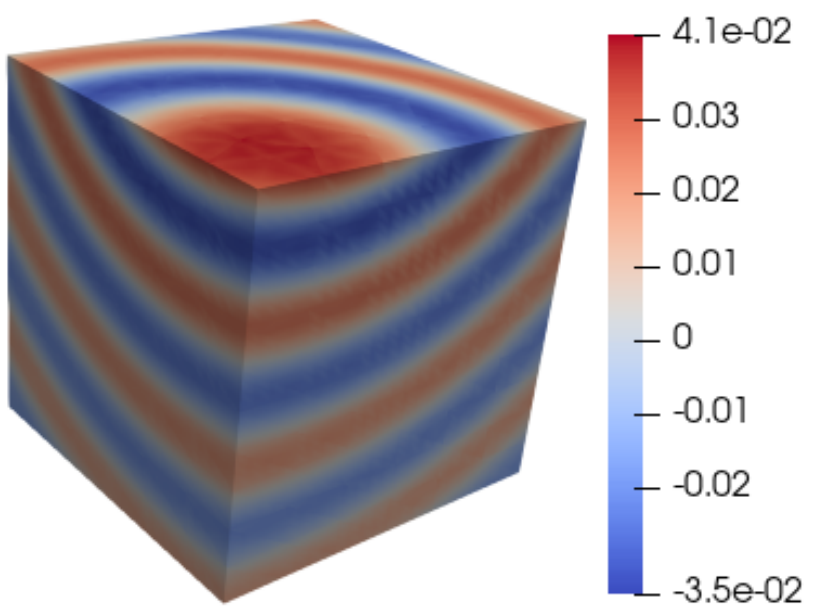


Code validation strategy









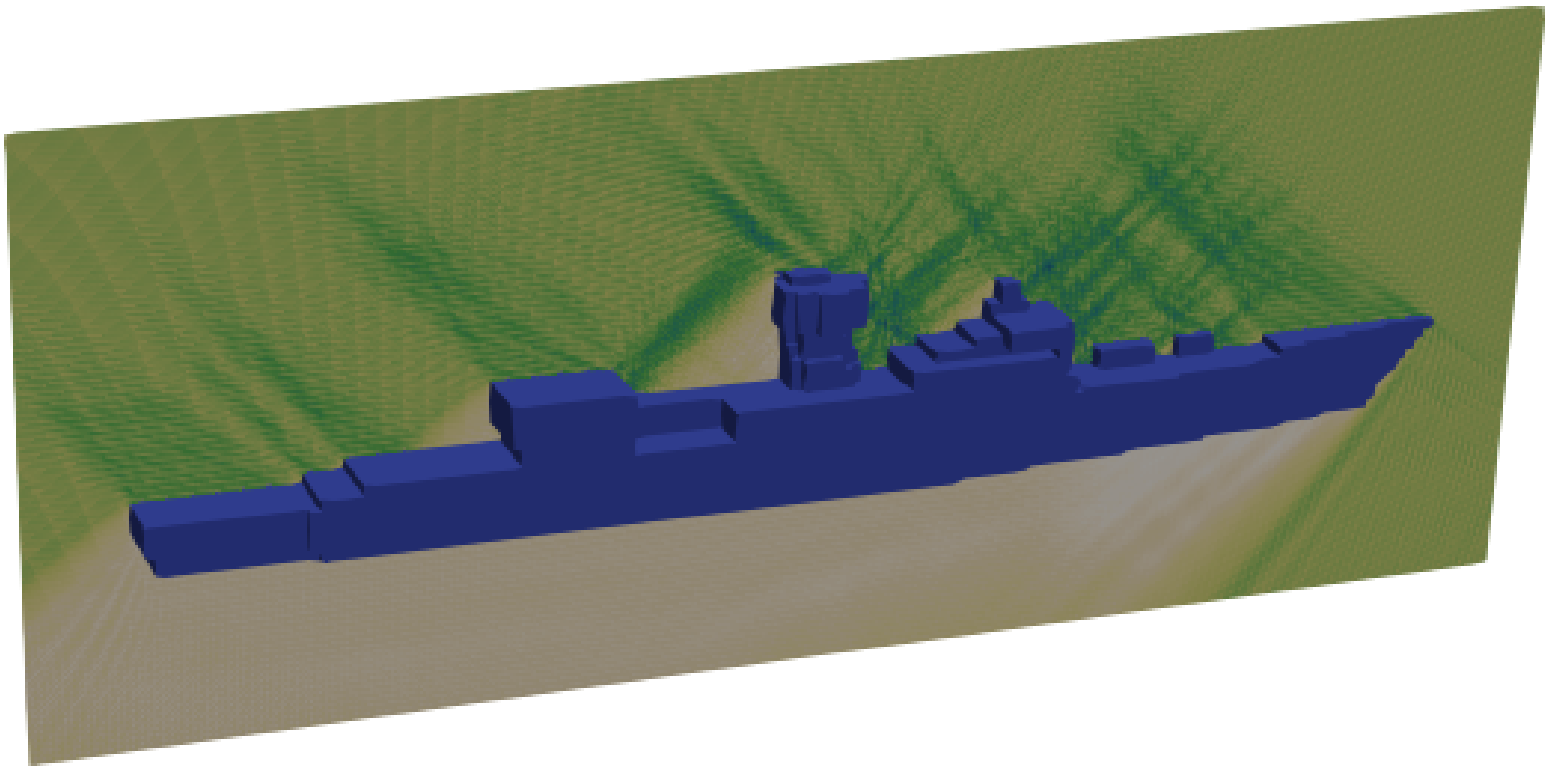
Scaling up

- Result taken from M. Sirdey PhD thesis (ONERA-Makutu) co-advised by S. Pernet (ONERA)
- Electromagnetic diffraction of a frigate illuminated from above

$$300 \times 100 \times 100 \ \lambda^3.$$

- Calculations involving more than a billion unknowns

Unknown	Matrix	LU Factorization	Method
16 Go	3 To	1000 To	800 Go



Conclusion

Itterative variational Trefftz methods are conducive to

- algebraic decomposition domain
- controling round-off error pollutions
- reducing dispersion errors
- HPC implementation

Some references

- Local strategies for improving the conditioning of the plane-wave Ultra-Weak Variational Formulation Hélène Barucq, Abderrahmane Bendali, Julien Diaz, Sébastien Tordeux Journal of Computational Physics, 2021, 441, pp.110449
- Ultra-weak variational formulation for heterogeneous maxwell problem in the context of high performance computing Sébastien Pernet, Margot Sirdey, Sébastien Tordeux 2022, Esaim Proc.
- A discontinuous Galerkin Trefftz type method for solving the two dimensional Maxwell equations Hakon Sem Fure, Sébastien Pernet, Margot Sirdey, Sébastien Tordeux SN Partial Differential Equations and Applications, 2020, 1 (23), pp.19.
- A Symmetric Trefftz-DG formulation based on a local boundary element method for the solution of the Helmholtz equation Hélène Barucq, Abderrahmane Bendali, M'Barek Fares, Vanessa Mattesi, Sébastien Tordeux Journal of Computational Physics, 2017, 330, pp.1069-1092.
- GoTem3 : Un code opensource fortran Trefftz Maxwell 3D.