



**HAL**  
open science

# Trefftz variational iterative methods for solving linear hyperbolic systems

Sébastien Tordeux

► **To cite this version:**

Sébastien Tordeux. Trefftz variational iterative methods for solving linear hyperbolic systems. Contemporary Challenges in Trefftz Methods, from Theory to Applications, Banff International Research station (Mexico), May 2024, Oaxaca, Mexico. hal-04637774

**HAL Id: hal-04637774**

**<https://hal.science/hal-04637774v1>**

Submitted on 3 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# 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 05/17



# Section 1

## Introduction

## Maxwell system

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

$$\begin{cases} \nabla \times \mathbf{H} &= ik \varepsilon_r \mathbf{E} \\ \nabla \times \mathbf{E} &= -ik \mu_r \mathbf{H}, \text{ dans } \Omega, \end{cases}$$

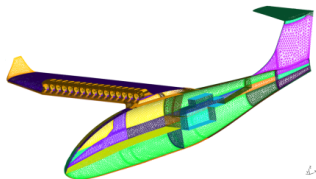
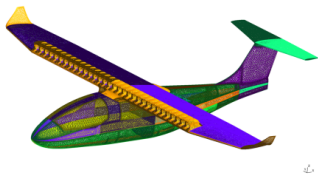
with boundary conditions

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

where  $\Omega$  is a 3D domain,  $\varepsilon_0$  and  $\mu_0$  are piecewise constant.

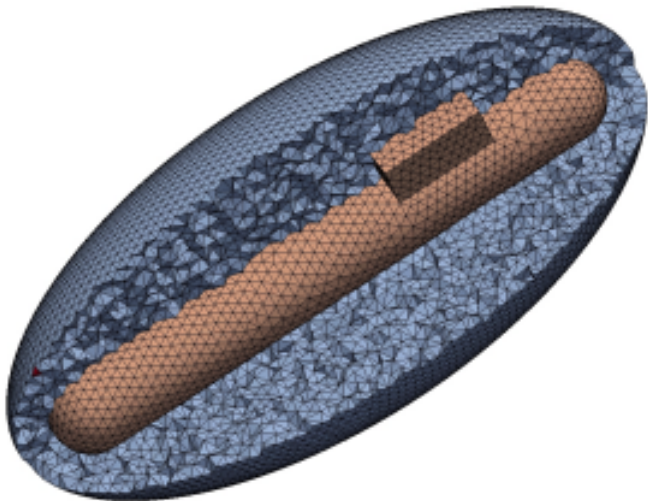
**Objective:** performing simulation in very large domains

→ **until 200 wavelengths =  $200\lambda = 8 \times 10^6$  elements.**



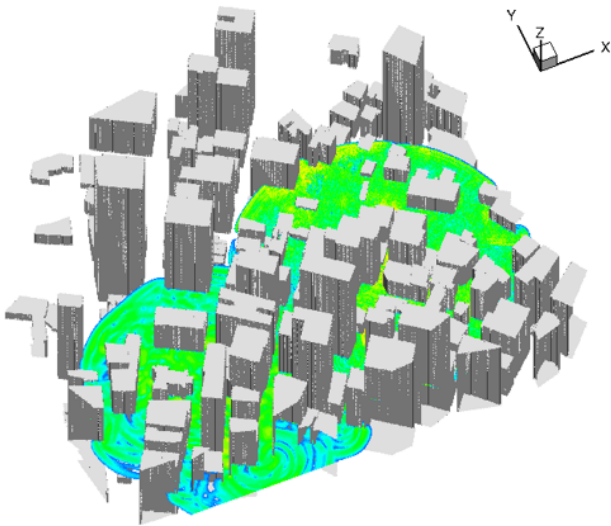
Geometry of Ampere prototype (DEMR ONERA).

## 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  $v \in H^1(\Omega)$

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

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 ?

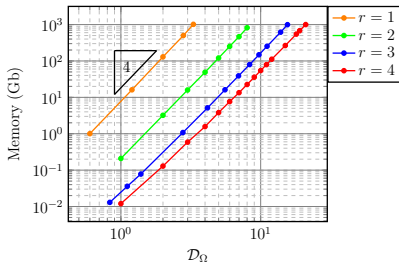
F. Ihlenburg and I. Babuska, *Finite element solution of the Helmholtz equation with high wave number part II: the hp version of the FEM*, *SIAM Journal on Numerical Analysis*, 34 (1), pp. 315–358 (1997).

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<sup>®</sup>) for different sizes of domain  $\mathcal{D}_\Omega$  and different approximation orders  $r$ .

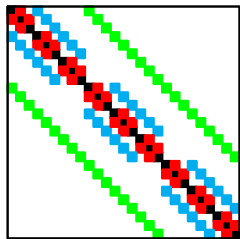
$$\text{Memory} = (\mathcal{D}_\Omega)^4.$$

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



# Matrix structure

$$AX = F$$



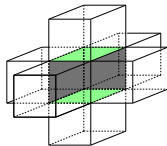
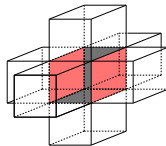
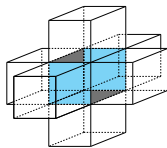
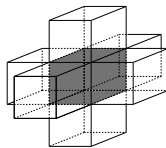
$A$

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

$X$

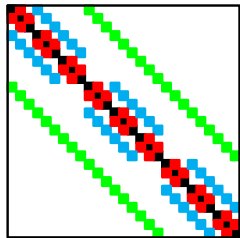
$=$

$F$

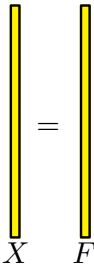


## Matrix structure

$$AX = F$$

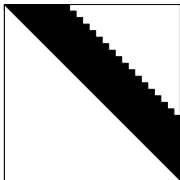
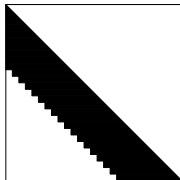


$A$



$$A = LU$$

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



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

	$L$				$U$
Size of the domain	$3\lambda$	$5\lambda$	$10\lambda$	$20\lambda$	$100\lambda$
Number of elements	27	512	1000	8000	$10^6$
LU Memory	10 MB	135 MB	4.3 GB	138 GB	<b>432 TB</b>

# 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  $\mathbf{A}\mathbf{X} = \mathbf{F}$

$$\mathbb{K}_k = \text{span}\left(\left\{\mathbf{F}, \mathbf{A}\mathbf{F}, \mathbf{A}^2\mathbf{F}, \dots, \mathbf{A}^{k-1}\mathbf{F}\right\}\right).$$

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

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

In theory, we have

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

$$\|\mathbf{F}\|_2 \geq \|\mathbf{A}\mathbf{X}_1 - \mathbf{F}\|_2 \geq \dots \geq \|\mathbf{A}\mathbf{X}_k - \mathbf{F}\|_2 \geq \|\mathbf{A}\mathbf{X}_{k+1} - \mathbf{F}\|_2$$

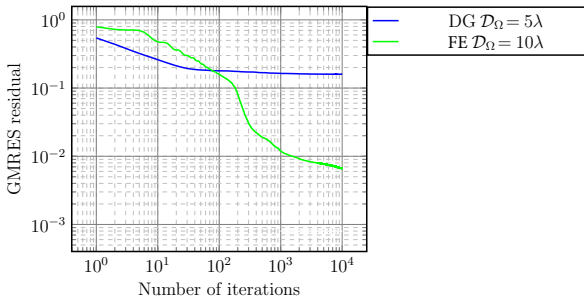
If the spectrum of  $A$  is included in  $B(\lambda_0, r)$  with  $r < |\lambda_0|$ , then  $\forall \varepsilon > 0, \exists C_\varepsilon > 0$

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

# GMRES in use

Efficient implementation suggested by Saad and Schultz<sup>1</sup>

- 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.

<sup>1</sup>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

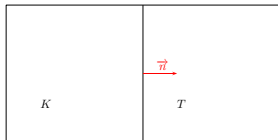
## – Global Preconditionnings

- 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

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

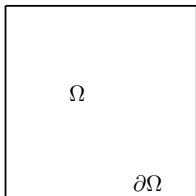
- A. Modave, X. Antoine et C. Geuzaine. An efficient domain decomposition method with cross-point treatment for Helmholtz problems. In CSMA 2019-14e Colloque National en Calcul des Structures. 2019.
  - F. Collino, P. Joly, and M. Lecouvez. "Exponentially convergent non overlapping domain decomposition methods for the Helmholtz equation." ESAIM: Math. Mod. and Num. An. 54.3. 775-810. 2020.
- ## – Hyperbolic Trefftz methods
- Back to hyperbolicity
  - 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  $\varphi \in L^2(\partial\Omega)$ .

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 $\Omega$

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

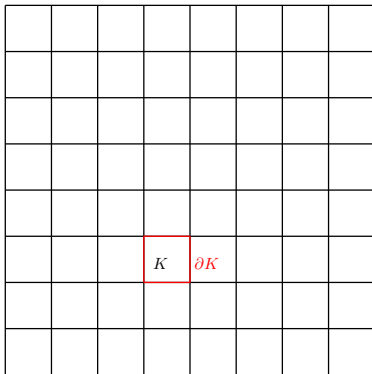


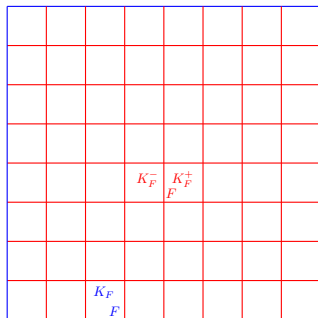
Figure: Domain discretization  $\Omega$ .



## Mesh of the domain $\Omega$

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

- $\mathcal{F}_K$  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**



$F \in \mathcal{F}_{ext}$   
 $F \in \mathcal{F}_{int}$

- 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:

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

We sum up all the elements:

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

We rewrite this expression in terms of jumps between elements:

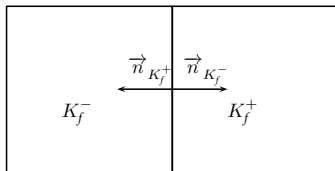
$$\sum_{F \in \mathcal{F}_{int}} \int_F \nabla p_F \cdot \overline{[p']_F} - p_F \overline{[\nabla p']_F} + \sum_{F \in \mathcal{F}_{ext}} \int_F \nabla p_F \cdot \overline{\vec{n}_{K_F}} \overline{p'_{K_F}} - p_F \overline{\nabla p'_{K_F} \cdot \overline{\vec{n}_{K_F}}} = 0,$$

assuming the solution is continuous on  $F$

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

The **jumps** are defined by

$$\begin{cases} \llbracket \vec{v} \rrbracket_F = \vec{v}_{K_F^+} \cdot \overline{\vec{n}_{K_F^+}} + \vec{v}_{K_F^-} \cdot \overline{\vec{n}_{K_F^-}}, \\ \llbracket p \rrbracket_F = p_{K_F^+} \overline{\vec{n}_{K_F^+}} + p_{K_F^-} \overline{\vec{n}_{K_F^-}}, \end{cases}$$



## Numerical traces

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

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

and for the exact solution, we have:

$$\widehat{\nabla p_F} = \nabla p_F \quad \text{and} \quad \widehat{p}_F = p_F$$

The numerical traces are defined by:

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

For the boundary condition, we have:

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

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

## The choice of coefficients $\alpha$ , $\beta$ , $\gamma$ and $\delta$

**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

$$\text{ingoing flux : } \quad \nabla p \cdot n - ikp,$$

$$\text{outgoing flux : } \quad \nabla p \cdot n + ikp.$$

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

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

It corresponds to:

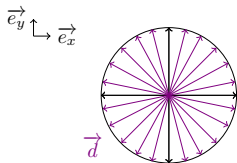
$$\alpha = 0 \text{ et } \beta = \frac{1}{2ik}.$$

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:

$$w_{d_n}(\mathbf{x}) = \exp\left(ikd_n \cdot \mathbf{x}\right),$$

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.

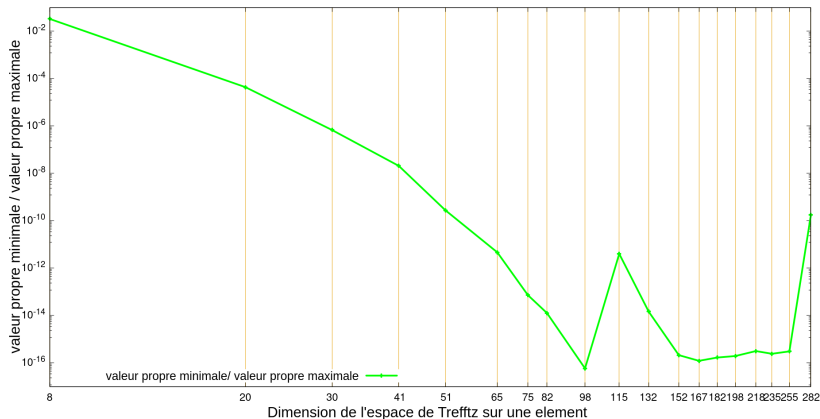
## Conditioning problem

Theoretically, plane waves form a **free family**.

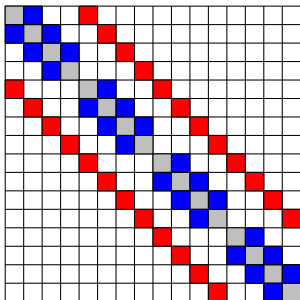
$$w_{d_n}(\mathbf{x}) = \exp(ikd_n \cdot \mathbf{x}),$$

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



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

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

- 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.

$$\left\{ \begin{array}{l} -i\omega \mathbf{v} = \nabla p \\ -i\omega p = \operatorname{div}(\mathbf{v}) \end{array} \right. \iff \left\{ \begin{array}{l} \frac{\partial \mathbf{v}}{\partial t} = \nabla p \quad \text{with } p(x, t) = p(x) \exp(-i\omega t), \\ \frac{\partial p}{\partial t} = \operatorname{div}(\mathbf{v}) \quad \text{with } \mathbf{v}(x, t) = \mathbf{v}(x) \exp(-i\omega t), \end{array} \right.$$

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

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

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

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

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

**Theorem:** Energy conservation formula

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

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

## First example of Friedrichs system: acoustics in 3D

$$\left. \begin{aligned} \rho \frac{\partial \mathbf{v}}{\partial t}(\mathbf{x}, t) + \nabla p(\mathbf{x}, t) &= 0, \\ \chi \frac{\partial p}{\partial t}(\mathbf{x}, t) + \operatorname{div} \mathbf{v}(\mathbf{x}, t) &= 0. \end{aligned} \right\} \iff \begin{cases} \rho \frac{\partial v_1}{\partial t}(\mathbf{x}, t) + \frac{\partial p}{\partial x_1}(\mathbf{x}, t) = 0, \\ \rho \frac{\partial v_2}{\partial t}(\mathbf{x}, t) + \frac{\partial p}{\partial x_2}(\mathbf{x}, t) = 0, \\ \rho \frac{\partial v_3}{\partial t}(\mathbf{x}, t) + \frac{\partial p}{\partial x_3}(\mathbf{x}, t) = 0, \\ \chi \frac{\partial p}{\partial t}(\mathbf{x}, t) + \frac{\partial v_1}{\partial x_1}(\mathbf{x}, t) + \frac{\partial v_2}{\partial x_2}(\mathbf{x}, t) + \frac{\partial v_3}{\partial x_3}(\mathbf{x}, t) = 0. \end{cases}$$

The unknown and the hyperbolic 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

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

- Total energy of the system:  $\int_{\Omega} \frac{Y^T M Y}{2} = \int_{\Omega} \frac{\rho |\mathbf{v}|^2 + \chi p^2}{2} dx$
- Work of external forces:  $\int_{\partial\Omega} \frac{Y^T F[n] Y}{2} = \int_{\partial\Omega} p \mathbf{v} \cdot \mathbf{n} ds_x$

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

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

The unknown and the hyperbolic matrices

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

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

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

## Is it really general ?

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

The matrix  $\mathbf{F}_i$  could 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

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

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:

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

## Riemann solver method in the heterogenous case

We start with

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

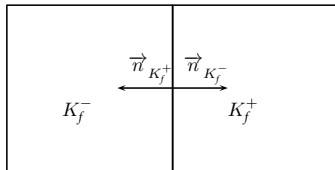
Then we focus on the faces and rewrite the previous formula as

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

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  $c$  ( $t = 0$ )

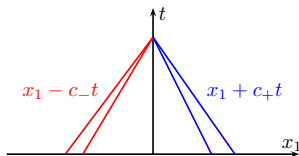
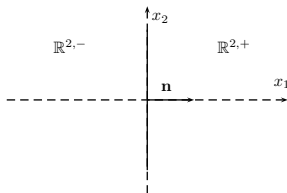
$$\begin{cases} Y(\mathbf{x}, 0) = Y_- & \text{pour } x_1 < 0, \\ Y(\mathbf{x}, 0) = Y_+ & \text{pour } x_1 > 0. \end{cases}$$

- Transmission conditions ( $x_1 = 0$ ).

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

Apply the characteristics method to get:

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



Condition initiale en  $t = 0$



## 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 boundary condition

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

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 functions by 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

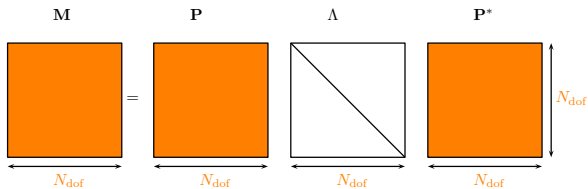
## Basis reduction strategy

**M** is **symmetrical**  
**positive definite**

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

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

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



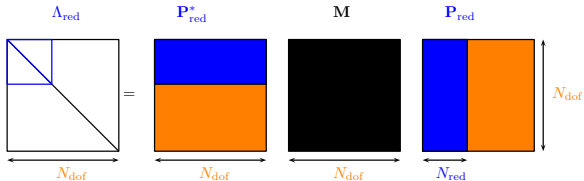
**P** is the orthogonal matrix formed with the eigenvectors.  
**Λ** is the matrix of eigenvalues

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

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{N_{\text{red}}} \\ \vdots \\ \lambda_{N_{\text{dof}}-1} \\ \lambda_{N_{\text{dof}}} \end{pmatrix} \rightarrow \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{N_{\text{red}}} \end{pmatrix}$$

**Filtering**

$$\frac{\lambda_i}{\lambda_1} < \varepsilon$$



## Normalized reduced basis

The discrete space is reduced:

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

à

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

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

The reduced basis can be normalized and the new matrix is

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

The diagram illustrates the reduced matrix  $\mathbf{A}_{red} = \mathbf{I}_{red} - \mathbf{N}_{red}$ . It shows a sequence of matrices: a square matrix  $\Lambda_{red}^{-\frac{1}{2}}$  (labeled  $\mathbf{I}_{red}$  in the equation), a blue rectangular matrix  $\mathbf{P}_{red}^*$ , a black square matrix  $\mathbf{A}$ , a blue rectangular matrix  $\mathbf{P}_{red}$ , and another square matrix  $\Lambda_{red}^{-\frac{1}{2}}$ . Dimensions are indicated by double-headed arrows below each matrix:  $N_{red}$  for the square matrices,  $N_{dof}$  for the black matrix  $\mathbf{A}$ , and  $N_{dof}$  for the blue matrix  $\mathbf{P}_{red}^*$  and  $N_{red}$  for the blue matrix  $\mathbf{P}_{red}$ .

# The reduction of the basis improves the memory

$N_{\text{red}}$  depends on the size of element  $K$  and on the truncation criterion  $\varepsilon$ .

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

**$0.25\lambda$  case**,  $N_{\text{kry}}^{\text{restart}} = 500$ .

$10^{-15}$	$\varepsilon$	$10^{-13}$	$10^{-11}$	$10^{-9}$	$10^{-7}$	$10^{-5}$	$10^{-4}$	$10^{-3}$	$10^{-2}$
<b>N = 196</b>	$N_{\text{red}}$	154	126	96	70	48	36	30	16
<b>1.57 Go</b>	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  
 $\implies$  Improvement of spectral properties

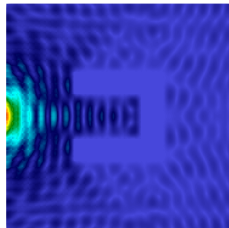


## How far can we reduce the basis?

$\varepsilon$	$10^{-7}$	$10^{-5}$	$10^{-4}$	$10^{-3}$	$10^{-2}$
Erreur	N/A	$3.4 \times 10^{-2}$	$9.8 \times 10^{-2}$	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

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

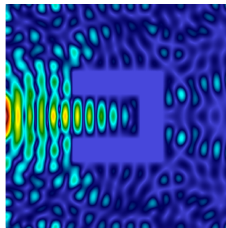


Electromagnetic Field Magnitude  
0.0e+00 4 6 8 1.1e+01



Good compromise

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

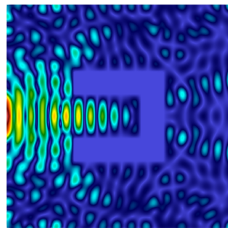


Electromagnetic Field Magnitude  
0.0e+00 4 6 8 1.1e+01



Reference solution

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



Electromagnetic Field Magnitude  
0.0e+00 4 6 8 1.1e+01



## 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

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

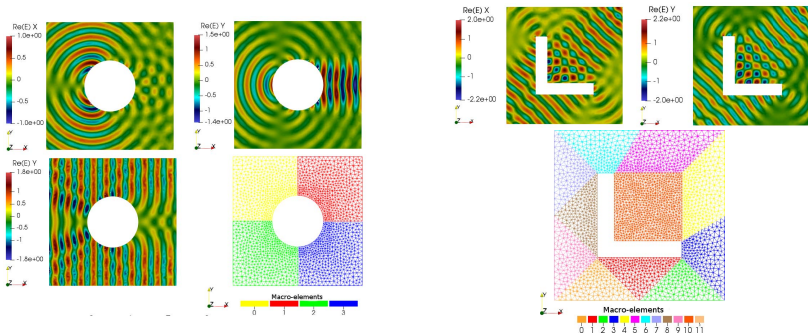
- A master Trefftz solver will couple the different solution methods (Previous formulation)

### Computational cost

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

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

## Two qualitative examples

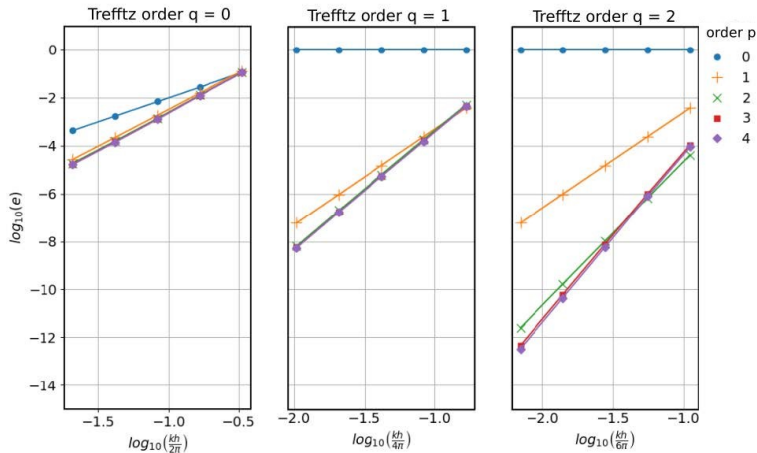


- Barucq, Bendali, Fares, Mattesi, Tordeux, A symmetric Trefftz-DG formulation based on a local boundary element method for the solution of the Helmholtz equation, JCP, (2017)
- Fure, Pernet, Sirdey, Tordeux, A discontinuous Galerkin Trefftz type method for solving the two dimensional Maxwell equations,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

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

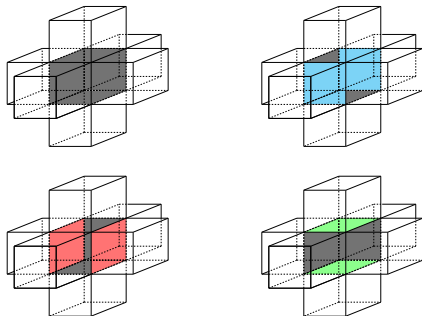
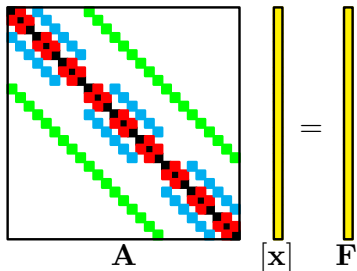
où

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

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

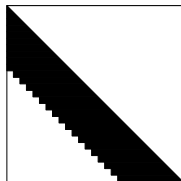
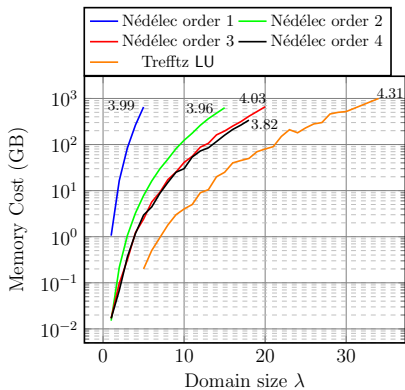
$$\text{taille}(\mathbf{A}) = N \times N_{\text{elem}}$$

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

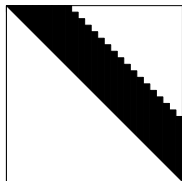


assembling of  $\mathbf{A}$  for  $N_{\text{elem}} = 27 = 3^3$  cubes.

## Memory limit of the direct solver



L

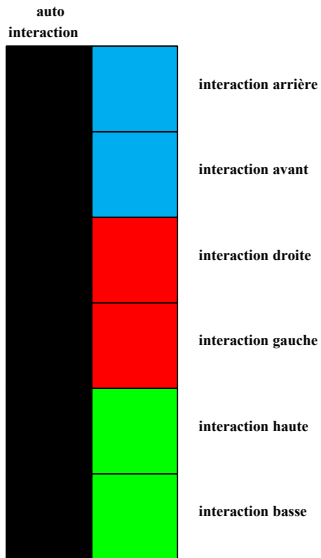


U

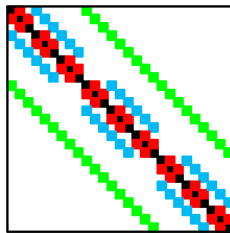
Domain size	$5\lambda$	$10\lambda$	$20\lambda$	$100\lambda$	$200\lambda$
Element number $N_{\text{elem}}$	512	1000	8000	$10^6$	$8 \times 10^6$
Memory of LU	268 Mo	4.3 Go	69 Go	43 To	688 To

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

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



Cartesian mesh  $\implies$  structured matrix



A

Without assembling,  $Ax$  can be computed for any  $x$



## Memory cost without matrix assembly

- $N_{\text{kry}}^{\text{restart}} = 50$  the dimension of the Krylov space
- $N_{\text{elem}}$  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**

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

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

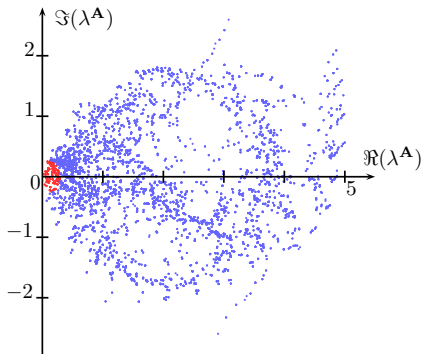
**Without assembling:** only the **interactions** are stored

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

# Cessenat-Després preconditioner

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

Idea: **préconditionning**  $A^\# A[x] = A^\# F$ .

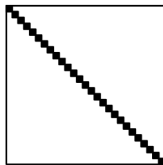


Real and imaginary parts of **A** spectrum, for a domain with size  $6\lambda$ .

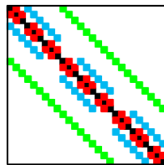
$$A = M + N$$

Cessenat-Després decomposition <sup>a</sup>

M



N



<sup>a</sup>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 **eigenvalues** and 0.

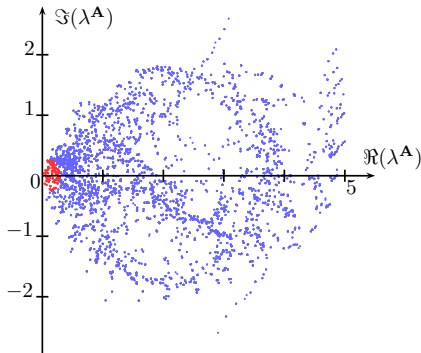
Idée:  $\mathbf{A}^\# \mathbf{A}[\mathbf{x}] = \mathbf{A}^\# \mathbf{F}$ .

$$\mathbf{A}^\# (\mathbf{M} + \mathbf{N})[\mathbf{x}] = \mathbf{A}^\# \mathbf{F}$$

$$\mathbf{A}^\# \mathbf{M}[\mathbf{x}]^{n+1} = \mathbf{A}^\# \mathbf{N}[\mathbf{x}]^n + \mathbf{A}^\# \mathbf{F}$$

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

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

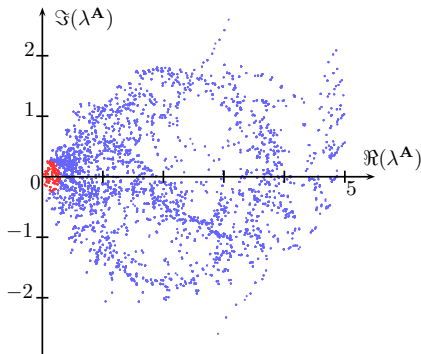


Real and imaginary parts of **A** spectrum, for a domain with size  $6\lambda$ .

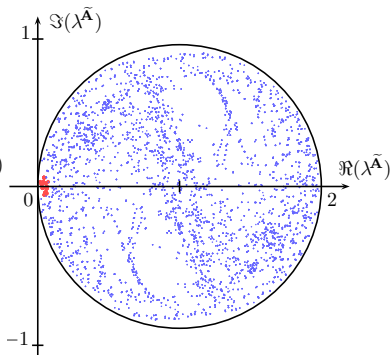
## Cessenat-Després preconditioner

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

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



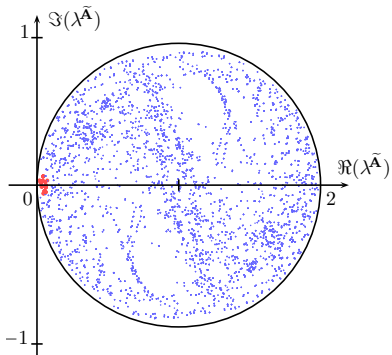
Real and imaginary parts of  $\mathbf{A}$  spectrum, for a domain with size  $6\lambda$ .



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

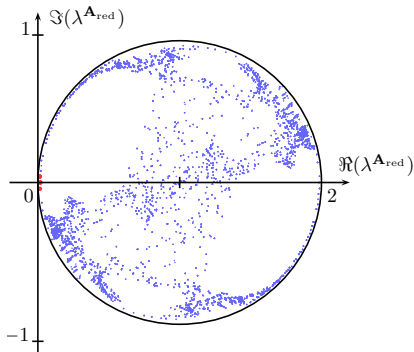
# Basis reduction improves matrix conditioning

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



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

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



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

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

Figure 1

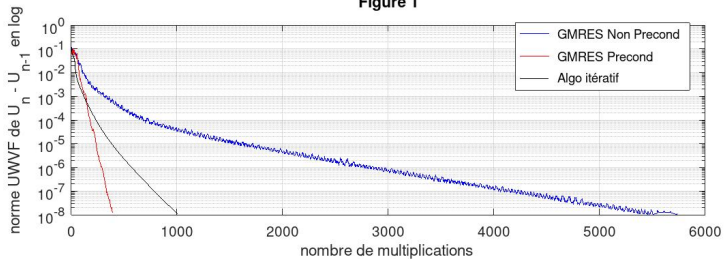
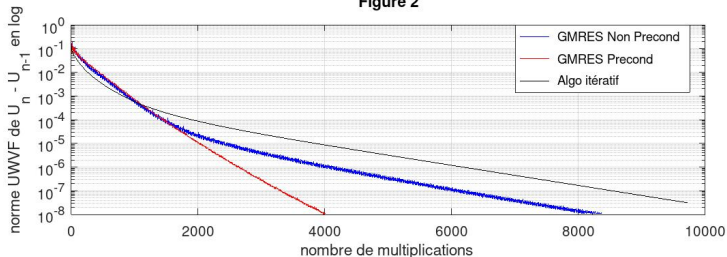
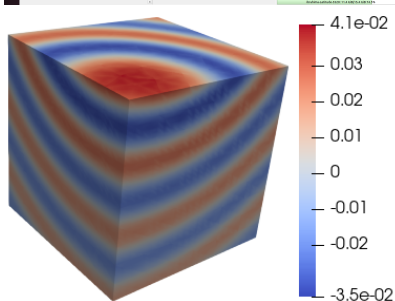
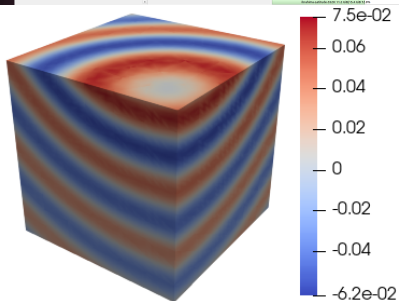
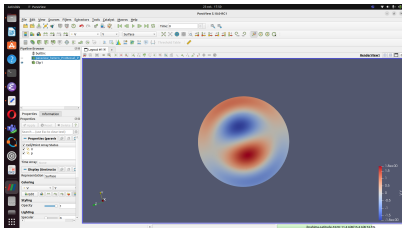
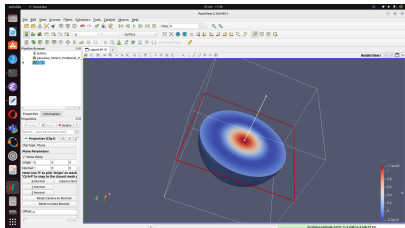


Figure 2



# Code validation strategy



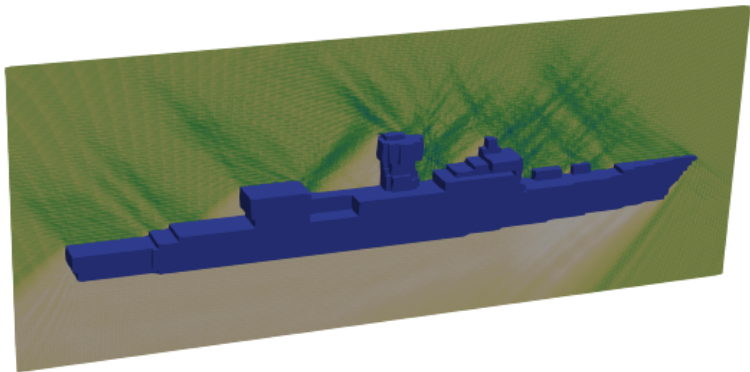
## Scaling up

- 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	<b>1000 To</b>	800 Go





# Conclusion

Iterative variational Trefftz methods are conducive to

- algebraic decomposition domain
- controlling 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 H el ene Barucq, Abderrahmane Bendali, Julien Diaz, S ebastien Tordeux Journal of Computational Physics, 2021, 441, pp.110449
- Ultra-weak variational formulation for heterogeneous maxwell problem in the context of high performance computing S ebastien Pernet, Margot Sirdey, S ebastien Tordeux 2022, Esaim Proc.
- A discontinuous Galerkin Trefftz type method for solving the two dimensional Maxwell equations Hakon Sem Fure, S ebastien Pernet, Margot Sirdey, S ebastien 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 H el ene Barucq, Abderrahmane Bendali, M'Barek Fares, Vanessa Mattesi, S ebastien Tordeux Journal of Computational Physics, 2017, 330, pp.1069-1092.
- GoTem3 : Un code opensource fortran Trefftz Maxwell 3D.